Mammographic Image Analysis: Risk Assessment and Microcalcification Classification Aspects

Zhili Chen

Department of Computer Science
Aberystwyth University

February 2013

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

This thesis has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date .....

Statement 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (candidate)

Date .....

Statement 2

I hereby give consent for my thesis, if accepted, to be made available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date .....

ABSTRACT

Breast cancer is currently the most common cancer among women worldwide. Mammography has been the most reliable and effective screening tool for the early detection of breast cancer. Recently, computer-aided diagnosis has become a major research topic in medical imaging and has been widely applied in clinical situations. This thesis investigates the employment of computer vision and image processing techniques for mammographic image analysis, which focuses on the aspects of mammographic risk assessment and microcalcification classification.
Many studies have indicated that mammographic density and mammographic parenchymal patterns are both strong predictive markers of breast cancer risk in mammographic images, which play an important role in estimating breast cancer risk. In this thesis, we present a variety of methods for estimating mammographic density and modelling mammographic parenchymal patterns. In order to build a complete framework for automated mammographic risk assessment, we first develop a breast region segmentation method as a pre-processing step, which segments the breast region from mammograms to provide fundamental data for subsequent analysis. Subsequently, we propose two breast density segmentation methods to estimate mammographic density. The first method is based on a modified fuzzy c-means algorithm which incorporates spatial information into the classic fuzzy c-means algorithm. The breast region is segmented into a number of sub-regions corresponding to different densities by clustering pixels with similar greylevel values. The second method exploits a topographic map to represent the overall profile of breast tissue density within the breast. Dense tissue regions are segmented by detecting *prominent/independent shapes* based on a *shape* tree. For modelling mammographic parenchymal patterns, we present a method to model breast tissue appearance based on statistical analysis of local tissue appearance. Five different types of local features are investigated, covering the aspects of intensity, texture, and geometric structure. In addition, a multiscale blob based representation is proposed to model mammographic parenchymal patterns. Instead of statistically describing breast tissue appearance within the whole breast, we focus only on dense tissue with approximately blob-like structures. The validity of the proposed methods is evaluated using the MIAS and DDSM databases. A high agreement with expert radiologists is indicated according to the BIRADS density classification. The obtained classification accuracies are up to 79.44% and 81.23%, and increase to 93.15% and 91.70% for the low/high risk classification.
For microcalcification classification, a novel method is developed based on topological analysis. The connectivity between individual microcalcifications is analysed to classify microcalcification clusters into malignant and benign. This method is evaluated using three datasets: MIAS, DDSM, and a full-field digital dataset. High classification accuracies (up to 96%) and good ROC results (area under the ROC curve up to 0.97) are achieved.

ACKNOWLEDGEMENTS

I am so grateful for the opportunity to undertake my PhD at Aberystwyth University, which has truly changed my life. I hesitated to do academic research in the past, but now I firmly believe that it has become an indispensable part of my life. Foremost, I would like to thank the person who made the biggest difference in my life, my supervisor, Prof. Reyer Zwiggelaar. Reyer is really a good teacher you will instantly like and never forget once you meet him. Without his guidance and encouragement, the work presented in this thesis would not have been possible. I often get lost in real life and cannot find my way home, but I have never lost myself in my research thanks to his wise directions and insightful suggestions. He is my primary first-aid station for receiving scientific knowledge and getting my problems solved during my research and writing of this thesis. I feel so blessed to have such a perfect supervisor, making my PhD study so meaningful and enjoyable. Over the past three years, not only have I gained invaluable research experiences from him, but I also learnt how to act as a good researcher and teacher, which will greatly benefit me for my whole life.

My special thanks also go to Dr. Frédéric Labrosse, for his helpful suggestions and thorough comments on my thesis draft. I would also like to acknowledge all members of the Vision, Graphics and Visualisation Group, for their feedback, questions, and suggestions on my research. I would also like to extend my heartfelt thanks to my colleagues in C57 (Alan, Bashar, Harry, Liping, Ran, Wenda, Yitian, and Ziming), for your self-giving help, for the days when we struggled together, and for all the fun we have shared. Many thanks also to the staff in the Department of Computer Science, for their friendliness and innumerable help.

My acknowledgement will never be complete without expressing my deepest gratitude to my family: Grandfather, Zuoxin Chen, Grandmother, Fenglan Huang, Father, Min Chen, Mother, Zhanping Xu, and Sister, Hongli Chen. I would not have been here without their constant love, support, and blessing. I also wish to thank my friends, for always being there in times of my needs. I believe that I am the luckiest person in the world to have such a nice family and so many friends, supporting me and loving me throughout my life.

Finally, I would like to thank the person who means the whole world to me, my husband, Yan Wu. Dear Tea Cup, thank you for always trusting me and encouraging me to follow my dreams. Without your encouragement, I would not have had the courage to start my PhD journey. Without your unconditional love and constant support, I would not have made it this far. Dear Tea Cup, thank you for all your understanding and tolerance, for all your care of my family and all your sacrifice in the last three years. Dear Tea Cup, I owe you a lot of gratitude, although you never want my thanks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Breast cancer is the most commonly diagnosed cancer in women all over the world, and it is the leading cause of cancer death among women worldwide (American Cancer Society, 2011b). Among different breast imaging modalities, mammography is still an effective diagnostic and screening tool for detecting breast cancer at its early stages. Although the effectiveness of mammography screening programmes in reducing breast cancer mortality still remains debated (Eurostat, 2009; Gøtzsche & Nielsen, 2009), numerous studies have shown that the early detection of breast cancer through mammography saves lives of women with breast cancer and increases treatment options (Nyström *et al.*, 1993; Tabár *et al.*, 2003; Berry *et al.*, 2005; Moss *et al.*, 2006; Kalager *et al.*, 2010; American Cancer Society, 2011b; Tabár *et al.*, 2011). Recently, computer-aided diagnosis has become a major research subject in medical imaging and has been widely applied in clinical radiology practice for assisting radiologists in diagnostic decision making. The scope of this thesis is mammographic image analysis using image processing and computer vision techniques, focusing on mammographic risk assessment and microcalcification classification aspects. In particular, mammographic risk assessment aims to estimate the risk of women developing breast cancer during their lifetime and classify mammographic images into different risk categories. Many studies have indicated that mammographic density and parenchymal patterns are strongly associated with breast cancer risk. Thus this thesis investigates different strategies for the quantitative analysis of mammographic density and the modelling of mammographic parenchymal patterns. On the other hand, microcalcification classification is intended to characterise and identify microcalcifications as malignant or benign. In this thesis, microcalcification clusters are analysed from a distinct viewpoint compared to existing approaches. Topological structures of microcalcification clusters, in terms of the connectivity between individual microcalcifications, are investigated for discriminating malignant from benign clusters. This chapter presents an overview of this thesis with an emphasis on the main

components, providing a brief outline of the research topics under investigation in this thesis. It starts with a brief introduction of breast cancer, which is followed by an explanation of mammography, introducing the acquisition process of mammograms and showing example mammograms acquired using the two most used viewpoints. On the basis of this, a concept of mammographic risk assessment is drawn and four well-known schemes for mammographic risk assessment are described. Subsequently, an introduction of microcalcifications in mammograms is given. After that, the use of computer-aided diagnosis in mammography and the clinical utility of the present research are discussed. Finally, aims and objectives of our research are summarised and an outline of this thesis is provided.

1.1 Breast Cancer Status

Breast cancer is a type of cancer originating in breast tissue, which is currently the most common cancer affecting women worldwide (American Cancer Society, 2011b). In European women, it is the leading cause of cancer death, causing 1 in 6 of all deaths from cancers (Eurostat, 2009). In the United States, a woman has a 12.15% (about 1 in 8) risk of developing breast cancer during her lifetime (American Cancer Society, 2011a). In the UK, breast cancer has been the most common cancer among women. It had the highest incidence rate among all cancers (124 cases per 100,000) during 2007-2009 (Office for National Statistics, 2012). In 2010, there were 49,564 new cases of breast cancer in women (157 cases per 100,000), accounting for 31% of all new cancer cases (Cancer Research UK, 2012a). Although breast cancer has a higher incidence rate in females compared to other types of cancers, lives of women with breast cancer can be saved by early detection. The five-year survival rate is as high as 98% among the US women diagnosed with breast cancer at an early stage (American Cancer Society, 2011b). In the European Union, the five-year survival rate is now around 80% (Eurostat, 2009). In fact, in most countries in the world, breast cancer survival rates have increased over the past decade, attributable to the improved cancer treatment and early detection through mammography (American Cancer Society, 2011b).

1.2 Mammography

Mammography is one of the most reliable and effective methods for detecting breast cancer at its early stages (Kolb *et al.*, 2002; Tabár *et al.*, 2003; Greif, 2010). In developed countries, population-based mammography screening programmes have been implemented (American Cancer Society, 2011b). Women are encouraged to participate in regular breast examinations through mammography. In

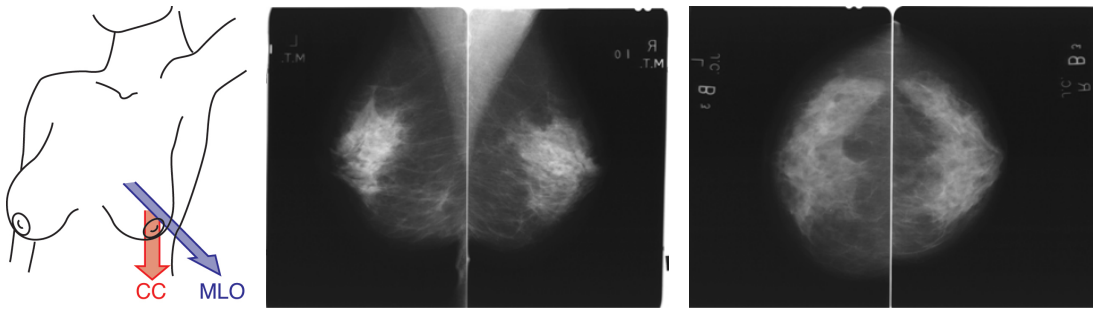
Figure 1.1: A film mammography system. The image was obtained with permission from http://www.medicalexpo.com/prod/siemens-ag-healthcare-sector/film-mammography-systems-70075-430251.html.

the United States, annual mammographic screening is recommended for women at normal risk, beginning at age 40 (National Comprehensive Cancer Network, 2012). In the UK, women aged between 50 and 70 years are invited for breast screening every three years. Recently, the age range has been extended to include women in their late 40s and up to 73 years (Cancer Research UK, 2012b).

Mammography is the process of using low-dose X-rays (usually around 30 kVp) to examine a breast and produce an X-ray image of breast tissue (i.e. mammogram), which is an effective diagnostic and screening tool for breast cancer in breast imaging. The spatial resolution of mammography is very high, normally in the range of $40 - 100\mu m$ per pixel. During the mammography procedure, the breast is compressed by a special mammography unit in order to reduce the thickness of tissue that X-rays must penetrate and therefore decrease the amount of scattered radiation and the required radiation dose. Parallel-plate compression is used to help hold the breast still and improve image quality (prevent motion blur). An example film mammography system is shown in Figure 1.1.

In mammographic screening, each breast is imaged separately with two views: the Medio-Lateral Oblique (MLO) view and the Cranio-Caudal (CC) view. Figure 1.2(a) shows the two viewpoints of X-rays. Figures 1.2(b) and 1.2(c) show two example mammograms of each view. For the same breast, each view is intended to show different appearances and details of breast tissue. Note that the pectoral muscle can be seen in the MLO view, which appears as a bright and triangular shaped region in the top right or left corner of the MLO mammograms.

Recently, there have been many improvements in mammography. One promising development is the adoption of digital techniques into conventional screen-



(a) Two viewpoints (b) MLO view (c) CC view

Figure 1.2: Two most used viewpoints for producing mammograms in mammography screening: (a) the direction of the two viewpoints (the image was taken from http://breastcancer.about.com/od/mammograms/a/mamm_views.htm); (b) the Medio-Lateral Oblique view; and (c) the Cranio-Caudal view.

film mammography. Full-field digital mammography enables the generation of an electronic image of the breast and therefore allows computer storage and operation. Some studies have compared full-field digital mammography with screen-film mammography in image quality, diagnostic accuracy, and cancer detection rate (Obenauer *et al.*, 2002; Pisano *et al.*, 2005; Skaane *et al.*, 2007; Vigeland *et al.*, 2008; Pisano *et al.*, 2008). In the study of Obenauer *et al.* (2002), it was reported that digital mammography was superior to screen-film mammography in image quality, detail visibility, image exposure, and artifacts. Pisano *et al.* (2005, 2008) compared the diagnostic accuracy of film versus digital mammography based on 49,528 women who underwent both film and digital mammography. As a means of screening, no significant difference was observed in the overall diagnostic accuracy between film and digital mammography (Pisano *et al.*, 2005). However, digital mammography appeared to benefit young women under the age of 50 and women with dense breasts, while film mammography nonsignificantly tended to improve the accuracy for women aged 65 years or older with fatty breasts (Pisano *et al.*, 2008). In another study based on women aged 45 to 69 years randomly assigned to film or digital screening mammography, digital mammography obtained a significantly higher cancer detection rate than that obtained by film mammography (Skaane *et al.*, 2007). However, in the study of Vigeland *et al.* (2008), no difference was seen between film and digital screening mammography for the detection of invasive cancers. The superior performance of full-field digital mammography to screen-film mammography as a screening tool still remains debated. The existing evidence is insufficient to support that digital mammography can replace screen-film mammography in routine mammography screening (National Comprehensive Cancer Network, 2012). In this thesis, digitised film mammograms are used for the development of mammographic image analysis methods.

1.3 Mammographic Risk Assessment

Mammographic risk assessment aims to estimate a woman's risk of developing breast cancer based on mammographic image analysis, which assists in identifying breast cancer at an early stage when treatment may be more effective. Mammographic parenchymal patterns and mammographic density are both strong predictive indicators of breast cancer risk, which are important mammographic features in the process of mammographic risk assessment. Many studies have indicated that an increased breast cancer risk is associated with mammographic parenchymal patterns showing high breast density (Wolfe, 1976a,b; Saftlas *et al.*, 1989; Boyd *et al.*, 1995, 1998, 2007; Harvey & Bovbjerg, 2004; Maskarinec *et al.*, 2005; Tabár *et al.*, 2005; McCormack & dos Santos Silva, 2006; Vachon *et al.*, 2007). Specifically, women with extremely dense breasts (the proportion of dense tissue is around 75% or larger than that) have a risk of developing breast cancer four to six times as high as the risk among women with fatty breasts.

Mammographic appearance of the breast varies among women due to differences in the tissue composition (the proportion of fatty, connective and epithelial tissue) and different X-ray attenuation properties of various tissues. Fatty tissue is radiographically lucent and would let most of X-rays pass through during the mammographic imaging process, and therefore appears dark in mammographic images. By contrast, non-fatty tissue (connective and epithelial tissue) is radiographically dense and would absorb more X-rays than fatty tissue during the acquisition process, and therefore appears brighter in mammographic images (Wolfe, 1976a; Boyd *et al.*, 1998, 2007). The different mammographic appearances generated by the variations in the tissue composition are referred to as mammographic parenchymal patterns, which represent the morphology, composition, and spatial distribution of various breast tissues. The bright appearance formed by non-fatty tissue is referred to as mammographic density, which reflects the relative amount of dense tissue in the whole breast.

Different methods for estimating mammographic risk have been presented based on mammographic parenchymal patterns and mammographic density. There are four main schemes used by radiologists for mammographic risk assessment: Wolfe, Boyd, BIRADS and Tabár, which can be divided into two groups according to whether mammographic parenchymal patterns or mammographic density plays a more significant role in assessing the risk of the development of breast cancer, although they interact with each other as indicators of breast cancer risk. The Wolfe and Tabár schemes assess mammographic risk mainly based on mammographic parenchymal patterns. By contrast, the Boyd and BIRADS schemes mainly rely on mammographic density. The details of these four schemes are described below.

1.3.1 Wolfe's Four Risk Categories

Wolfe (1976a,b) proposed four categories of mammographic risk with respect to different parenchymal patterns. The classification is as follows:

- N1 (lowest risk): Parenchyma is composed primarily of fat with at most small amounts of "dysplasia", and no ducts are visible.
- P1 (low risk): Parenchyma is chiefly fat with prominent ducts in anterior portion up to one-fourth of the volume of the breast, and also may be a thin band of ducts extending into a quadrant.
- P2 (high risk): Severe involvement with prominent duct pattern, occupying more than one-fourth of the volume of the breast.
- DY (highest risk): Severe involvement with "dysplasia", often obscuring an underlying prominent duct pattern.

1.3.2 Boyd's Six Class Categories

Boyd *et al.* (1995) classified mammograms into six categories according to the extent of mammographic density, referred to as Six Class Categories (SCC). The percentage of the breast occupied by dense tissue increases from SCC1 to SCC6, and this increasing mammographic density indicates an increased risk of breast cancer. The percentages of dense tissue of the six categories are as follows:

- SCC1: None
- SCC2: $< 10\%$
- SCC3: $[10\% - 25\%\rangle$
- SCC4: $[25\% - 50\%\rangle$
- SCC5: $[50\% - 75\%\rangle$
- SCC6: $\geq 75\%$

1.3.3 Four BIRADS Density Categories

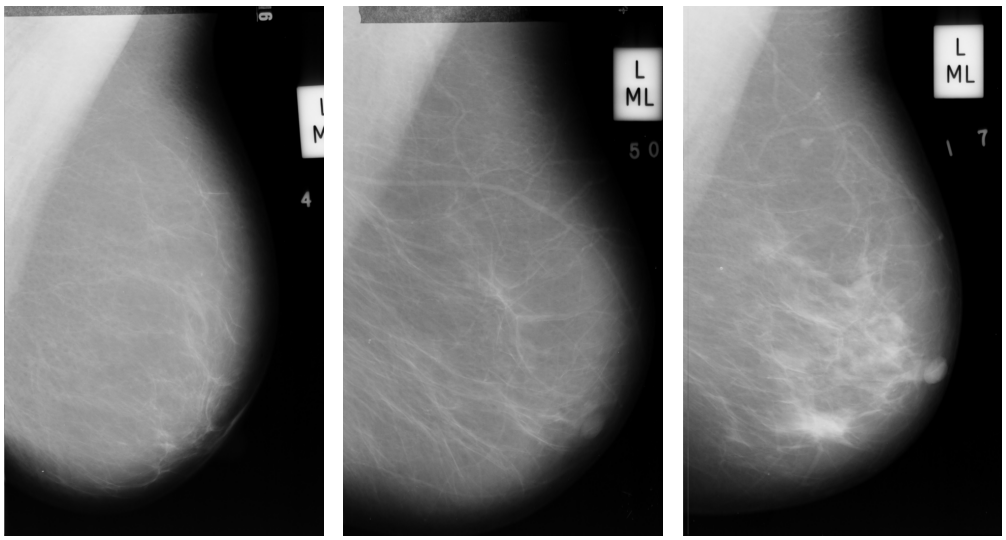
The American College of Radiology (1998) developed the Breast Imaging Reporting and Data System (BIRADS) where BIRADS density classification was included to inform the referring physicians of the decline in sensitivity of mammography with the increase in breast density. There are four BIRADS density categories corresponding to an increasing breast density, which are defined as follows:

- BIRADS I: The breast is almost entirely fatty.
- BIRADS II: There is some fibroglandular tissue.
- BIRADS III: The breast is heterogeneously dense.
- BIRADS IV: The breast is extremely dense.

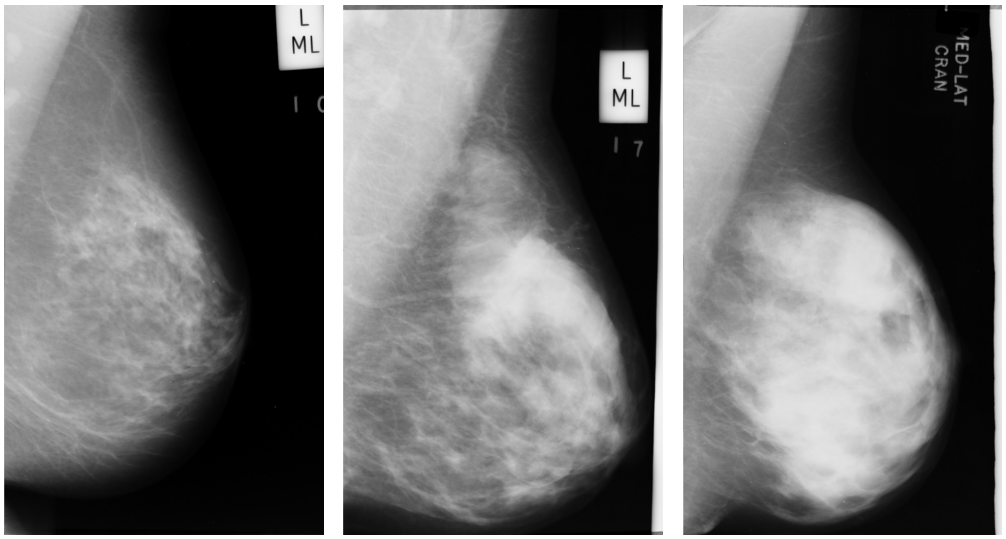
1.3.4 Tabár's Five Patterns

Tabár *et al.* (2005) proposed a model of breast tissue in mammographic images, containing four mammographic building blocks: nodular densities, linear densities, homogeneous densities, and radiolucent areas. These four building blocks constitute the normal anatomy of the breast. In particular, nodular densities mainly correspond to terminal ductal lobular units (TDLU) which visually look like small bright blobs; linear densities correspond to either ducts, fibrous strands or blood vessels; homogeneous densities correspond to fibrous tissue which could obscure the underlying normal TDLU and ducts as well as their alterations due to hyperplastic breast changes; and radiolucent areas represent adipose tissue, which appear as dark areas. According to the composition of the four mammographic building blocks, Tabár *et al.* (2005) defined five mammographic parenchymal patterns. Patterns I to III refer to low risk, while Patterns IV and V refer to high risk. The relative compositions of the four building blocks for the five Patterns I to V are as follows:

- Pattern I: $[25\%, 15\%, 35\%, 25\%]$
- Pattern II: $[2\%, 14\%, 2\%, 82\%]$
- Pattern III: similar to Pattern II
- Pattern IV: $[49\%, 19\%, 15\%, 17\%]$
- Pattern V: $[2\%, 2\%, 89\%, 7\%]$



(a) mdb255ll (b) mdb273ll (c) mdb083ll



(d) mdb007ll (e) mdb115ll (f) mdb067ll

Figure 1.3: Example mammograms covering different categories of the four schemes for mammographic risk assessment: (a) N1, SCC1, BIRADS I, Pattern II; (b) N1, SCC2, BIRADS I, Pattern II; (c) P1, SCC3, BIRADS II, Pattern III; (d) P1, SCC4, BIRADS II, Pattern I; (e) P2, SCC5, BIRADS III, Pattern IV; and (f) DY, SCC6, BIRADS IV, Pattern V.

Figure 1.3 shows example mammograms covering all categories of the four schemes: Figures 1.3 (a)/(b), (c)/(d), (e), and (f) correspond to Wolfe's four categories (N1, P1, P2, and DY); Figures 1.3 (a)-(f) correspond to Boyd's six categories (SCC1-6); Figures 1.3 (a)/(b), (c)/(d), (e), and (f) correspond to the four BIRADS categories (BIRADS I-IV); and Figures 1.3 (d), (a)/(b), (c), (e), and (f) correspond to Tabár's five patterns (Pattern I-V).

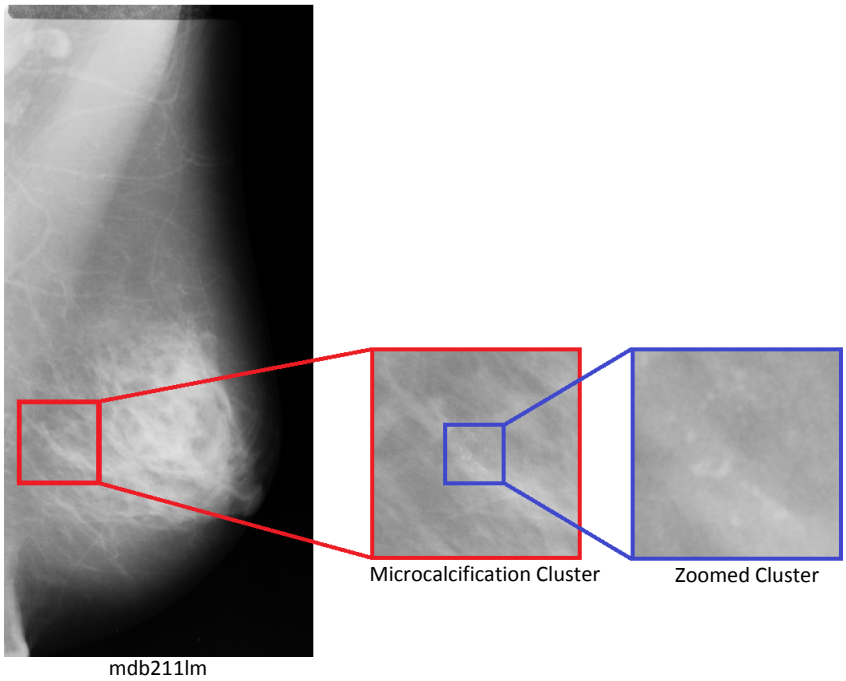


Figure 1.4: An example microcalcification cluster. The area containing the microcalcification cluster is cropped from the mammogram. The cluster region in the image patch is zoomed to improve the visibility of microcalcifications.

1.4 Microcalcifications in Mammograms

Microcalcifications are small deposits of calcium in breast tissue that appear as small bright spots in mammograms (Shen *et al.*, 1994; Dhawan *et al.*, 1996; Cheng *et al.*, 2003; Ma *et al.*, 2010), and a group of microcalcifications closely located in an area form a microcalcification cluster. The radiological definition of microcalcification clusters is that at least three microcalcifications are present within a 1 cm^2 region (Ma *et al.*, 2010; Soltanian-Zadeh *et al.*, 2004). Due to its high spatial resolution ($40 - 100\mu m$ per pixel), mammography enables the detection of microcalcifications at an early stage. Figure 1.4 shows an example microcalcification cluster present in a mammogram.

The presence of microcalcification clusters in mammograms is a primary sign of breast cancer. However, not all microcalcification clusters necessarily indicate the presence of cancer, only certain kinds of microcalcifications are associated with a high probability of malignancy (Sickles, 1986). It is difficult for radiologists to distinguish malignant from benign cases, which results in a high rate of unnecessary biopsy examinations (Cheng *et al.*, 2003; Soltanian-Zadeh *et al.*, 2004). The diagnostic gold standard of microcalcification clusters (benign or malignant) in this thesis has been provided by biopsy examination.

1.5 CAD in Mammography

Computer-aided diagnosis (CAD) has become one of the major research topics in medical imaging and diagnostic radiology (Doi, 2007). CAD systems are designed to assist radiologists in making the diagnostic decision in clinical situations. Radiologists use the computer output as a "second opinion" and make the final decision. As mentioned in Section 1.2, mammography is an effective screening and diagnosis tool for breast cancer. However, it is time consuming for radiologists to read/analyse considerable numbers of mammograms taken in screening programmes, and inaccurate risk assessments may be made by even the most experienced radiologists due to the long hours of work. Moreover, there is a substantial intra/inter-observer variability in radiologists' interpretations of the same mammograms (Berg *et al.*, 2000; Ooms *et al.*, 2007). For instance, using four BIRADS density categories, only moderate agreement ($\kappa = 0.43$) was shown across five experienced observers in (Berg *et al.*, 2000). Ciatto *et al.* (2005) also reported that only moderate agreement ($\kappa = 0.54$) was seen between twelve observers on the four-category scale and average intra-observer agreement was 0.71. For the detection/diagnosis of breast cancer through mammography, abnormalities and cancers may be missed (false negatives), and non-cancerous lesions may be misdiagnosed as cancers (false positives) (Nishikawa, 2007). Therefore, it is necessary to develop CAD using image processing and computer vision techniques for assisting radiologists in the interpretation of mammograms and improving the diagnostic consistency and accuracy.

A large number of CAD systems have been employed for assisting radiologists in screening mammography (Doi, 2007; Nishikawa, 2007). Warren Burhenne *et al.* (2000) investigated the capability of CAD in identifying the missed lesions by radiologists in a retrospective study of 427 prior mammograms. It was shown that CAD could potentially reduce the false negative rate while retaining the recall rate (no increase in the recall rate was indicated). This claim was subsequently confirmed by the study of Freer & Ulissey (2001) in which 12,860 mammograms were interpreted with and without the assistance of CAD. It was indicated that the use of CAD could increase the detection rate of early-stage cancers without excessive effect on the recall rate or positive predictive value for biopsy. On the other side, different findings were provided by some studies. Gur *et al.* (2004) assessed changes in mammography recall and breast cancer detection rates after introducing a CAD system into a clinical radiology practice in an academic setting. 115,571 screening mammograms were used in this study: 59,139 were interpreted by radiologists with the assistance of the CAD system, while 56,432 were interpreted without the use of CAD. No statistically significant changes in the recall and detection rates

were found after the introduction of CAD into this clinical practice due to its level of performance at that stage. Consequently, it is difficult to measure the clinical benefits of the use of CAD due to variability in breast cancer cases present in the population-based screening over years. However, recent clinical studies have indicated that using CAD can help radiologists detect more cancers which would otherwise be missed and the use of CAD can improve radiologists' efficiency in the interpretation process of mammogrpahic images (Nishikawa, 2007). In fact, numerous computer-based mammographic image analysis methods have been developed for CAD schemes. A literature review of computerised mammographic image analysis will be provided in Chapter 2, focusing on the methods related to automated mammographic risk assessment. In addition, a brief review of various methods for characterising and classifying microcalcifications in mammograms will be provided in Chapter 7.

1.6 Clinical Utility of the Present Research

The present research mainly covers two aspects of mammographic image analysis: mammographic risk assessment and microcalcification classification. As indicated in Section 1.3, mammographic risk assessment is intended to assess a woman's risk of developing breast cancer through mammography, which is expected to play an important role in mammographic screening and early detection of breast cancer. Women at high risk of breast cancer may be recommended more intensive surveillance, use of breast MRI as an adjunct to mammography, or implementation of prevention strategies, such as reducing known risk factors as much as possible, having surgery or chemoprevention (Amir *et al.*, 2010; National Comprehensive Cancer Network, 2012). Initial methods for assessing mammographic risk were entirely subjective, performed by radiologists according to qualitative assessment schemes (Yaffe, 2008). The lack of objective and quantitative methods resulted in significant inconsistent assessment results between and within radiologists (Harvey & Bovbjerg, 2004). Only moderate inter-observer agreement was reported when using the BIRADS breast density scheme (Berg *et al.*, 2000; Ciatto *et al.*, 2005). It is difficult for radiologists to make both accurate and consistent risk assessment for the mammograms generated in mammographic screening. In addition, as indicated in Section 1.4, the presence of microcalcification clusters is an important sign of early breast cancer. However, it is usually difficult for radiologists to discriminate malignant from benign microcalcification clusters due to the variability associated with their appearances (Cheng *et al.*, 2003). This results in false positives or false negatives in malignancy diagnosis of microcalcifications. Many computer-aided mammographic image analysis methods have been developed in

the past decade, which enables the development of breast cancer screening and diagnosis towards more accurate and objective directions. Our research concentrates on investigating and developing a series of mammographic image analysis methods to achieve automated, objective, accurate, and reliable mammographic risk assessment and microcalcification classification in clinical practice. It is deemed to contribute to computer-aided diagnosis of breast cancer and benefit women undergoing mammography.

1.7 Research Aims and Objectives

As mentioned above, mammographic density and mammographic parenchymal patterns are both strong predictive indicators of breast cancer risk in mammographic images, which play an important role in mammographic risk assessment. Our research aims to analyse mammographic images using image processing and computer vision techniques, and develop a set of methods for automated mammographic risk assessment. We propose to estimate mammographic density by means of breast density segmentation. The breast is segmented into different types of tissue with different densities, or only dense tissue is extracted from the breast. As such, the resulting tissue composition (the proportions of different densities) or the resulting percent area of dense tissue in the breast can be used for mammographic risk assessment. On the other hand, we plan to model mammographic parenchymal patterns based on statistical texture analysis of mammographic tissue appearance. The resulting models in the form of occurrence histograms can be used for mammographic risk assessment. We also attempt to create a new representation of mammographic parenchymal patterns in contrast to using texture features, which can provide a quantitative description of mammographic density and can be used for mammographic risk assessment. Figure 1.5 shows the schema of our proposed framework for automated mammographic risk assessment.
In addition, we propose to develop an automated microcalcification classification method for distinguishing malignant from benign cases. This method is distinct from existing approaches which concentrate on the morphology of individual microcalcifications and/or global cluster features.
In summary, the main aims of our research are as follows:

Our research focuses on mammographic image analysis using image processing and computer vision techniques and aims to develop a new framework for automated mammographic risk assessment as well as a new method for the classification of microcalcification clusters in mammograms.

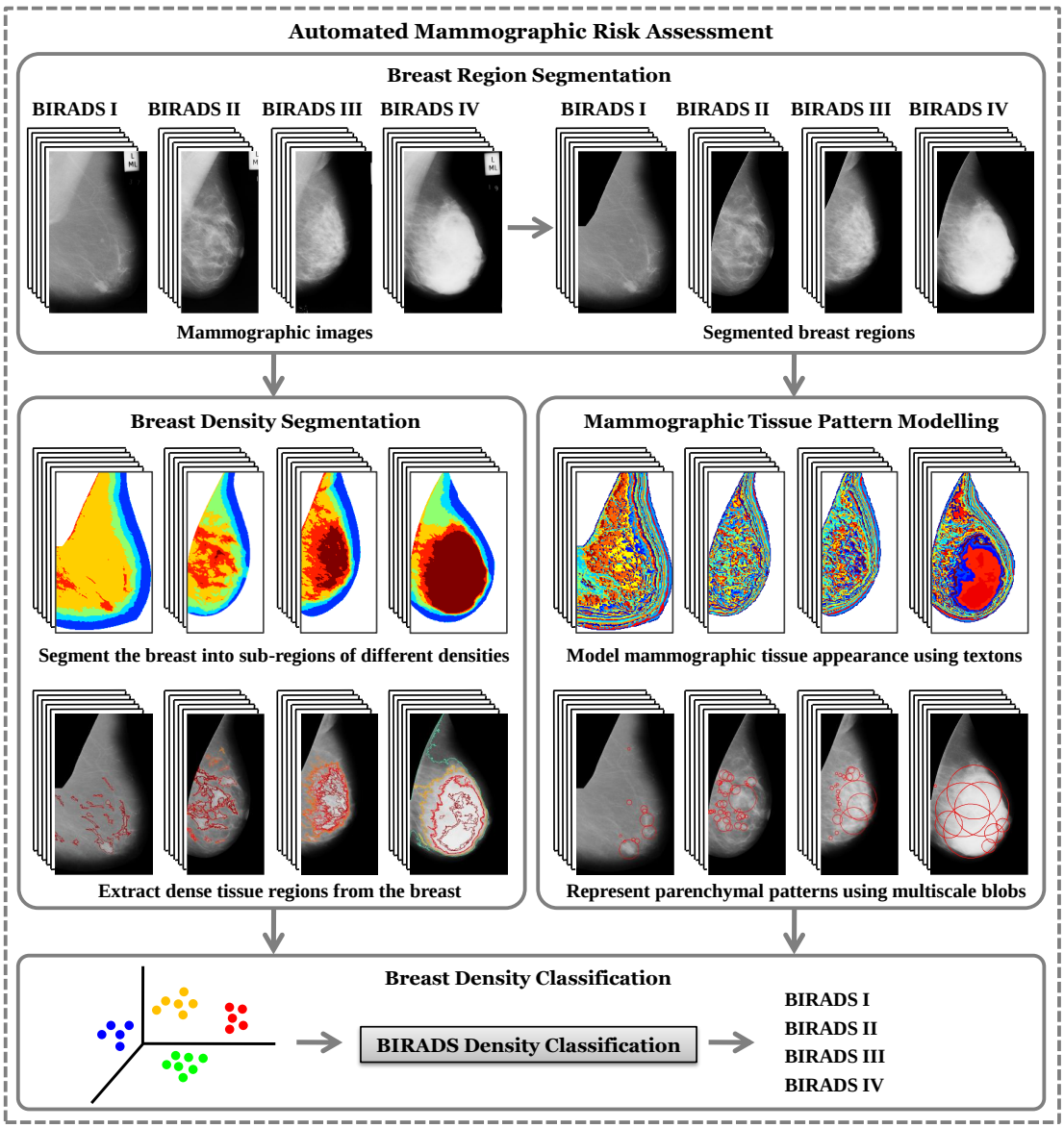


Figure 1.5: The schema of automated mammographic risk assessment.

We subdivide the main goals of this thesis into a set of objectives described below:

- The development of a method for breast region segmentation. This method can separate the breast from the background and remove the pectoral muscle in MLO mammograms. This is an essential pre-processing step which identifies the region of interest (i.e. breast region) for subsequent analysis.

- The development of two methods for breast density segmentation. The first method can segment the whole breast region into a number of sub-regions which correspond to different types of tissue with different densities. The second method concentrates on dense tissue within the breast and can detect dense tissue regions from the breast.

- The investigation of modelling mammographic parenchymal patterns based on statistical analysis of local tissue appearance. A visual dictionary is generated to summarise local tissue appearance with descriptive "words". The overall parenchymal patterns in mammographic images are represented as occurrence histograms over this dictionary.

- The creation of a new representation of mammographic parenchymal patterns by focusing on approximately blob-like tissue patterns. Bright blob-like structures are detected over a range of scales, and mammographic parenchymal patterns are represented with a set of multiscale blobs.

- The application of the developed methods to automated mammographic risk assessment using benchmark databases for mammographic image analysis.

- The development of a new method for classifying microcalcification clusters in mammograms into malignant and benign cases.

1.8 Thesis Outline

The remaining chapters of this thesis are organised as follows:

- In Chapter 2, a literature review of existing approaches to mammographic image analysis is provided, including mammographic image segmentation, mammographic density estimation, mammographic parenchymal patterns characterisation, and breast density classification.

- In Chapter 3, a brief review of the most used image segmentation methods is provided, where the core concepts of the various approaches are introduced and their main advantages and disadvantages are discussed. On the basis of this, our developed mammographic image segmentation methods are presented, including one breast region segmentation method and two breast density segmentation methods.

- In Chapter 4, a variety of texture analysis approaches are reviewed, focusing on the different strategies for extracting texture features. Then, a number of texture representations based on local features are investigated. On the basis of this, an approach to modelling mammographic parenchymal patterns is proposed, where different types of features are extracted from local neighbourhoods and are statistically analysed to build an overall model of mammographic tissue appearance.

- In Chapter 5, an introduction of scale-space theory is first provided to describe a general idea of multiscale image representation. Subsequently, a number of blob detectors are briefly reviewed. Finally, a multiscale blob based representation is proposed, which describes mammographic parenchymal patterns focusing on blob-like dense tissue regions.

- In Chapter 6, all the methods proposed in Chapters 3-5 are evaluated for the purpose of mammographic risk assessment. The details of experimental data, evaluation methodologies, and experimental results are provided.

- In Chapter 7, a method for the classification of microcalcification clusters in mammograms is proposed. A brief review of related work is first provided, followed by a basic introduction to mathematical morphology. Subsequently, the details of the proposed method are described. The experimental results and a full comparison are provided at the end of this chapter.

- In Chapter 8, a comprehensive analysis of the obtained results is presented, including an internal comparison across our proposed methods and a complete comparison with closely related publications. The performance of these approaches and possible directions for further work are discussed.

- In Chapter 9, conclusions and contributions of this thesis are summarised.

Chapter 2

A Literature Review of Mammographic Image Analysis

In this chapter, we present a review of (semi-)automatic analysis of digitised film mammograms or full-field digital mammograms in the literature. We trace the development history of computerised analysis of mammographic images back to the 1990s and present a brief taxonomy of existing approaches. Clearly, it is impossible to include all existing approaches in this chapter. We here focus on publications related to our research and provide a brief description of the main points of these approaches. We begin by presenting an overview of mammographic image segmentation in Section 2.1. Firstly, a set of methods for breast region segmentation are introduced, which is an essential pre-processing stage in mammographic image analysis. Subsequently, various breast density segmentation methods are outlined, which aim to separate dense tissue from fatty tissue or segment the entire breast region into multiple sub-regions of different densities. In Section 2.2, we present a number of methods for estimating breast density, which are mainly divided into two groups: the first uses percent breast density based on the derived dense tissue from the previous step of breast density segmentation; and the second uses volumetric breast density based on the estimation of the thickness of dense tissue. In Section 2.3, a review of methods for characterising mammographic parenchymal patterns is provided, focusing on different types of features extracted for discriminating parenchymal patterns of women at different risk levels of developing breast cancer. In Section 2.4, we concentrate on the methods in particular for breast density classification where models of breast tissue are built for classifying mammographic images according to breast density categories. We conclude this chapter in Section 2.5 by making a summary of this literature review, and discuss how it establishes a basis for our research and leads to our own work that will be presented in the following chapters.

2.1 Mammographic Image Segmentation

2.1.1 Breast Region Segmentation

Breast region segmentation is an essential pre-requisite in the (semi-)automatic analysis of (digital or digitised) mammographic images, which aims to separate the breast region from background information. It normally consists of two independent segmentations, which are breast-background segmentation and pectoral muscle segmentation, respectively. The first identifies the boundary between the breast and background which usually contains annotations and labels, and the second identifies the boundary of the pectoral muscle (present in MLO views). A precise segmentation of the breast region with the minimum loss of breast tissue is important for further mammographic image analysis, such as detecting abnormalities, modelling parenchymal patterns, and accurate registration.

2.1.1.1 Breast Boundary Identification

There are two major problems in the accurate identification of the breast boundary. The first problem is that the region near the breast boundary has decreasing contrast, caused by a lack of uniform compression of breast tissue during the acquisition process (Ojala *et al.*, 2001; Wirth & Stapinski, 2003). For digitised mammographic images, the digitisation process may even further decrease the visibility of the breast boundary due to additional noise. The low visibility of the breast skin line makes it difficult to identify the breast boundary. The other problem is the non-uniform background that may contain high-intensity regions, such as annotations, labels, frames, and unexposed regions, which interferes with the segmentation of the breast region (Ojala *et al.*, 2001; Wirth & Stapinski, 2003). There have been various approaches to the breast boundary identification in mammograms. The traditional methods are based on histogram thresholding and gradient analysis. Lau & Bischof (1991) as well as Yin *et al.* (1994) used a simple thresholding to segment the background from the breast region and they respectively used average filtering and morphological filtering to obtain a smooth boundary. Bick *et al.* (1995) combined thresholding with region growing and morphological filtering. Heine *et al.* (1998) proposed an approach based on multiresolution wavelets to separate the breast region from the background. A local norm image was calculated from one of the wavelet expansion images which was afterwards thresholded to eliminate the background noise. The breast region was located using a region labelling routine. Karssemeijer (1998) applied a global thresholding technique to segment breast tissue from the background. The threshold value was automatically determined by performing peak detection on the histogram. Masek

et al. (2000) proposed a local thresholding method to solve the problem caused by the non-uniform background. Raba *et al.* (2005) used multiple thresholds computed from a smoothed histogram to obtain a set of overlapping masks. A final threshold was determined by statistically analysing the greylevel distribution of the region defined by the boundaries of the smallest and the largest masks, which was subsequently used to threshold the image. After that, a connected component labelling algorithm was applied to extract the largest region as the breast region. Méndez *et al.* (1996) applied a gradient based algorithm to detect the breast edge. As a pre-processing step, a two-level thresholding was used to obtain an approximate breast region. An average filter was applied to smooth the entire mammogram. Then, five reference points were automatically selected to divide the breast into three regions. Finally, a tracking algorithm was used to search for edge pixels using greylevel gradients. The obtained results were quantitatively evaluated and compared with those of Yin *et al.* (1994). Abdel-Mottaleb *et al.* (1996) used histogram thresholding to segment the interior region of the breast. The remaining region near the breast boundary was located by an edge detection technique. The union of the two resulting regions formed the final breast region. Morton *et al.* (1996) proposed another gradient based method. First, the background was estimated by an initial threshold determined by the largest background peak, an edge was then detected by a horizontal line-by-line gradient analysis. Zhou *et al.* (2004) modified the above method to improve its performance. After the preliminary edge was detected by the gradient analysis, a more accurate tracking algorithm was applied to search for new edge points. Finally, a simple linear interpolation was used to find a continuous breast boundary. One of the inherent limitations of the methods stated above is that very few of them can entirely preserve the breast skin line and nipple. One promising method for the breast boundary identification was developed by modelling the background as a polynomial (Chandrasekhar & Attikiouzel, 1996, 2000). This method was based on the assumption that the intensity values of the background are spatially continuous, lower than the breast, and within a closed interval. The image was first thresholded to approximately separate the background region which contained the entire background and a small portion of the breast adjoining it. A polynomial was fitted to all the pixels within the resulting region in order to model the background. The background was then subtracted from the image using the polynomial model. Finally, the resulting image was thresholded and further processed to yield a binary labelled image with a smooth border, representing the breast region.

On the other hand, due to the fact that the breast boundary is a well defined curve, some methods have applied active contours to segment the breast region. Ojala *et al.* (2001) incorporated active contours for the segmentation of the breast

region. A global thresholding was used for preliminary segmentation. Morphological filtering was applied to define the boundary of the resulting region after thresholding, which produced a non-smooth boundary. Three contour modelling approaches, Fourier descriptors, snakes and spline functions, were investigated to smooth the boundary. The snake approach provided the best results and the spline approach performed worst. Ferrari *et al.* (2000, 2004b) developed two methods for the identification of the breast boundary. The two methods started with logarithmic contrast enhancement. A binarisation procedure using the Lloyd-Max algorithm (Lloyd, 1982) was then applied. A morphological opening operator was used to remove spurious details within the resulting binary image, and an approximate breast boundary was extracted using the chain-code algorithm (Gonzalez & Woods, 2002). After that, in (Ferrari *et al.*, 2000), the greylevel information of the pixels along the normal lines to the approximate breast boundary was analysed to estimate the skin-air interface of the breast. Finally, a traditional snake model was applied to detect the breast boundary. In (Ferrari *et al.*, 2004b), the approximate breast boundary was used as the input to an adaptive active contour model, which included a balloon force and an adaptive gradient force in the energy formulation. Wirth & Stapinski (2003) also applied active contours to solve the problem of the segmentation of the breast region. Preliminary thresholding, edge enhancement, and noise removal were first performed to generate a suitable input for the active contour stage. A modified greedy snake algorithm was used to generate the final breast boundary.

In addition, the breast region segmentation can be considered as a classification problem in which the task is to classify the image pixels into the breast or non-breast region. As such, some methods for the segmentation of the breast region are based on pixel-wise classification using pixel-based features. Lou *et al.* (2000) used a cluster classification method which was based on an adaptation of the c-means algorithm (Bezdek, 1981) to obtain an initial boundary. For each initial boundary point, a normal line with the direction towards the background was plotted, and the pixels on each normal line were traced to locate the skin-air boundary point. These extrapolated boundary points were then refined and smoothed to generate the final breast boundary. Saha *et al.* (2001) used scale-based fuzzy connectivity for segmenting the breast region from the background. The pixels with connectivity strengths greater than half the maximum strength were classified as the background and discarded, while the pixels with lower connectivity strengths were retained as the breast region. Wirth *et al.* (2005) proposed a rule-based fuzzy reasoning algorithm. A pre-processing step was performed based on attribute morphology to suppress artifacts and increase the homogeneity of the mammogram. Pixel classification measures were computed and used as the input of a fuzzy de-

Table 2.1: A summary of existing approaches to breast boundary identification, indicating authors, years of publication, techniques, experimental data for test, and segmentation performance.

Author & Year	Thresholding	Gradient	Modelling	Active Contours	Classification	Test Data	Segmentation Performance
Lau & Bischof (1991)	√					20	not mentioned
Yin *et al.* (1994)	√					80	not mentioned
Bick *et al.* (1995)	√					740	97% acceptable
Méndez *et al.* (1996)		√				156	89% (nearly) accurate
Abdel-Mottaleb *et al.* (1996)	√	√				500	98% acceptable
Morton *et al.* (1996)	√	√				1000	$>$ 90% accurate
Chandrasekhar & Attikiouzel (1996)	√		√			58	57 sufficient
Heine *et al.* (1998)	√					42	87% true
Karssemeijer (1998)	√					615	not mentioned
Masek *et al.* (2000)	√					322 MIAS	75% no obvious defect
Chandrasekhar & Attikiouzel (2000)	√		√			32 MIAS	94% acceptable
Ferrari *et al.* (2000)	√			√		66 MIAS	61 accurately detected
Lou *et al.* (2000)					√	422	421 faithfully extracted
Ojala *et al.* (2001)	√		√	√		20	90% acceptable
Saha *et al.* (2001)					√	120	100% correct
Wirth & Stapinski (2003)	√			√		25 MIAS	1.6% and 1.1% error*
Zhou *et al.* (2004)	√	√				744	89.78% very close
Ferrari *et al.* (2004b)	√			√		84 MIAS	FP = 0.41%, FN = 0.58%⋆
Raba *et al.* (2005)	√					322 MIAS	98% nearly accurate
Wirth *et al.* (2005)					√	322 MIAS 44 DDSM	0.996/0.981† 1.0/0.93 (24), 0.99/0.95 (20)‡

The numbers outlined in the column of Test Data indicate the number of images (mammograms) used for evaluation in the approaches. MIAS and DDSM are two well-known mammogram databases which will be introduced in Chapter 6.

* 1.6% error percentage on average for the breast region and 1.1% for the background (pixel-based quantitative analysis).

⋆ the average false positive (FP) and false negative (FN) rates were 0.41% and 0.58% (pixel-based quantitative analysis).

† the mean values of the completeness and correctness measures were 0.996 and 0.981, which indicated the proportion of the ground truth detected by the algorithm and the percentage of correctly detected breast region, respectively.

‡ the 44 mammograms from DDSM contain 24 normal cases and 20 benign cases. The mean values of the completeness and correctness measures were 1.0 and 0.93 for the 24 normal mammograms, and 0.99 and 0.95 for the 20 benign mammograms.

cision making process to classify the mammogram into the background and the breast region. A summary of all the approaches above can be found in Table 2.1.

2.1.1.2 Pectoral Muscle Identification

The pectoral muscle appears as a bright and triangular shaped region in the top right or left corner of MLO view mammograms. Automatic identification of the pectoral muscle is useful in many aspects of mammographic image analysis. First of all, the presence of the pectoral muscle can cause a large number of false positives when detecting suspicious masses, as it may have a similar appearance to that of mammographic parenchyma. Similarly, the inclusion of the pectoral muscle in the image data being processed for breast density quantification can affect the performance of intensity-based methods, since it appears to have approximately the same intensity as dense tissue in the image. Hence, the exclusion of the pectoral muscle is necessary to avoid bias in the detection of breast cancer and automatic breast density quantification. In addition, the segmented pectoral muscle can be used in 3D reconstruction from multiple views of mammograms and it is also an important landmark for mammogram registration and comparison.

The initial methods for automatic identification of the pectoral muscle were based on the Hough transform. It was assumed that the boundary of the pectoral muscle would be an approximately straight line and should be located in the upper part of the image. As the first step of these algorithms, a region of interest (ROI) including the pectoral muscle was always marked using this prior knowledge. Karssemeijer (1998) first extracted a triangular ROI and applied the Sobel operator to calculate the gradient within the ROI. A Hough transform was then performed on the gradient image. The Hough space can be considered as an accumulator array of all possible straight lines. Each peak in the Hough space corresponds to a straight line in the image. Peaks were detected based on a sliding window. The points which were not a local maximum within the window or which were smaller than a threshold value were set to zero. After that, only one peak was selected from the remaining candidate points according to the area of the pectoral muscle segmented by the corresponding line in the image. Finally, the selected peak in the Hough space was inversely transformed into the image space and all pixels above the line in the ROI were marked as the pectoral muscle. Ferrari *et al.* (2000) modified the method proposed by Karssemeijer (1998). They used geometric and anatomical constraints to eliminate unlikely pectoral lines instead of using the threshold value in the procedure of peak selection. First of all, an approximate ROI was located automatically by six control points obtained based on the breast contour. After that, a Gaussian filter was used to remove noise in the ROI. The Hough transform

was then applied to the Sobel gradient of the ROI. To detect the true pectoral muscle edge, a filtering procedure was performed to remove the lines intercepting the top of the image outside the ROI or with slopes outside the possible range. Furthermore, each remaining Hough accumulator cell was weighted to enhance the Hough transform peaks. Finally, the Hough accumulator cell with the maximum value was detected to represent the pectoral muscle edge.

To overcome the limitation of the hypothesis of a straight-line representation of the pectoral muscle, Ferrari *et al.* (2004a) proposed a new method based on a multi-resolution technique using Gabor wavelets. Firstly, a Gabor filter bank was designed to enhance the directional piece-wise linear structures in the ROI containing the pectoral muscle. Subsequently, the ROI was convolved with the Gabor filter bank, and the magnitude and phase components of the filtered images were combined by vector summation to compute the magnitude and phase images. The magnitude value of each pixel in the ROI was propagated in the phase direction until encountering an edge-flow coming from an opposite direction. The candidate edges were then detected from the resulting image. Finally, a post-processing step was used to identify the true pectoral muscle boundary by connecting disjoint boundary segments and removing false edges. It was indicated that the Gabor filter based method provided more consistent results than the method based on the Hough transform. In the same year, Kwok *et al.* (2004) also proposed a new method consisting of two major components. First, the pectoral muscle boundary was approximated with a straight line, which was then refined to a curve using an iterative cliff detection. The straight line was estimated within the ROI based on iterative threshold selection and straight line fitting incorporating a gradient test. The resulting line was validated to adjust the ROI and perform a second estimation. The new straight line was then used to initialise the cliff detection, where the intensity profiles along a set of search paths perpendicular to the straight line were analysed and the cliff locations were detected by fitting the smoothed profile to a sigmoid function. This procedure was performed iteratively to obtain a more accurate curve. Finally, a closed region was generated to segment the pectoral muscle. This method was adaptive to large variations in edge strength and texture of the pectoral boundary. In addition, it remained effective when the pectoral boundary was partially obscured by overlaid glandular tissue or artifacts.

In contrast to the edge based methods, Raba *et al.* (2005) proposed a method based on region growing. The breast orientation was first identified by detecting the curvature on each side of the breast profile. A seed point was then placed at the first pixel of the non-curved side. The region growing started from this seed point to fill the pectoral muscle based on an intensity criterion. A constraint on the size of the grown region was applied to avoid an overgrown region.

In addition to incorporating general shape and location assumptions of the pectoral muscle, most of the above methods only focus on local information of the image. Ma *et al.* (2007) used two graph theory based segmentation algorithms, where global information of the image was incorporated by means of identifying a region based on properties of the whole region instead of those near its boundary only. One is the adaptive pyramid (AP) algorithm and the other is the minimum spanning tree (MST) algorithm. The AP algorithm is a hierarchical approach, where the original image is regarded as a graph, and a sequence of graphs derived from this base graph build a pyramid graph, which forms a multi-resolution representation. Each surviving vertex in the highest level graph represents one component of the original image. The MST algorithm starts with a graph representing the original image where all vertices (i.e. pixels) are regarded as separate components. In subsequent iterations, two disjoint components are merged if the external variation between the two components is small compared to the internal variation within both components. The disjoint components in the final graph form the segmentation. The AP and MST algorithms were performed to identify the general location and shape of the pectoral muscle. The initial segmentation results were used to initialise an active contour procedure to obtain the smooth pectoral muscle boundary. The AP and MST algorithms followed by active contours as described above did not significantly outperform the Gabor filter based method proposed by Ferrari *et al.* (2004a). However, the AP algorithm did outperform previous methods in terms of the number of images with small error. Moreover, it was the first attempt to use graph theory based methods to find the pectoral muscle in mammograms. Ma *et al.* (2006) evaluated the two graph theoretic segmentation algorithms. Four radiologists were requested to draw the pectoral muscle boundary independently for each mammogram. The simultaneous truth and performance level estimation method (Warfield *et al.*, 2004) was used to estimate the true boundary and the sensitivity and specificity of the two segmentation algorithms, based on the four manually drawn boundaries and the two automatically detected boundaries. The results showed that the performance of the AP algorithm was not statistically different from the radiologists, which seemed to claim that the current best methods for the identification of the pectoral muscle were near the inherent limitations of this task.

Recently, Camilus *et al.* (2010) proposed a method based on graph cuts. Firstly, a rectangular ROI including the whole pectoral muscle was extracted. The graph cut based segmentation technique was then applied to segment the ROI and identify the pectoral muscle edge. The segmentation started with a weighted graph in which each vertex was considered as a group. In each iteration of the segmentation, two adjacent groups were merged if the merging criterion was satisfied. Finally, the

Table 2.2: A summary of existing approaches to pectoral muscle identification, indicating authors, years of publication, techniques, experimental data for test, and segmentation performance. HT, ED, CF, RG, and GT stand for Hough transform, edge detection, curve fitting, region growing, and graph theory, respectively.

Author & Year	HT	ED	CF	RG	GT	Test Data	Segmentation Performance
Karssemeijer (1998)	√	√				615	571 a pectoral muscle was segmented
Ferrari *et al.* (2000)	√	√				84 MIAS	FP = $1.98 \pm 6.09\%$; FN = $25.19 \pm 19.14\%$†
Ferrari *et al.* (2004a)		√				84 MIAS	FP = $0.58 \pm 4.11\%$; FN = $5.77 \pm 4.83\%$†
Kwok *et al.* (2004)		√	√			322 MIAS	83.9% adequate
Raba *et al.* (2005)				√		322 MIAS	86% good
Ma *et al.* (2006)					√	82	STAPLE, sensitivity score/specificity score‡
Ma *et al.* (2007)					√	84	80% average error less than 2mm
Camilus *et al.* (2010)			√		√	84	FP = 0.64%; FN = 5.58%†

† FP: false positive; FN: false negative (pixel-based quantitative analysis).
‡ STAPLE: simultaneous truth and performance level estimation.

identified pectoral muscle edge was refined using Bezier curve modelling. It was demonstrated by the validation results that the performance of this method was clearly better than the AP and MST algorithms, and its segmentation accuracy was comparable with the Gabor filter based method. Moreover, with respect to the number of images with small error, this method outperformed existing methods including the Hough transform, the Gabor filter, the AP and MST algorithms. A summary of these approaches is provided in Table 2.2.

As described above, many methods for breast region segmentation have been developed based on various image processing techniques. However, it is still a difficult task to achieve accurate segmentation of the breast region in particular for digitised mammograms, due to the presence of severe noise and the non-uniform background. Therefore, in order to implement a complete framework for mammographic risk assessment, we decided to develop a breast region segmentation method to obtain the region of interest to be analysed from each mammogram, by combining a variety of image segmentation techniques. Effectively, we intend to use the advantages of the various techniques simultaneously to achieve an improved overall performance compared to using these techniques individually. A further review of the well-known image segmentation methods will be presented in Section 3.1 where their main advantages and disadvantages are discussed, and a new method for breast region segmentation will be proposed in Section 3.2.

2.1.2 Breast Density Segmentation

As mentioned in Section 1.3, many studies have indicated that breast density is a strong risk factor for breast cancer (Boyd *et al.*, 1995, 1998, 2007; Harvey & Bovbjerg, 2004; Maskarinec *et al.*, 2005; McCormack & dos Santos Silva, 2006;

Vachon *et al.*, 2007), and therefore quantitative analysis of breast density is of great importance for mammographic risk assessment. Breast density segmentation aims to segment the breast based on density characteristics, which plays an important role in estimating breast density. Many of the segmentation methods segment dense tissue from the breast region and the area of the segmented dense tissue can be used for the estimation of breast density, which is based on an assumption that the breast is mainly composed of fatty and dense tissue. On the other hand, some segmentation methods segment the entire breast region into a number of sub-regions corresponding to a range of densities. The relative proportions of the resulting sub-regions of different densities can be used to estimate the overall density of the breast. In addition, breast density segmentation can be considered as an initial step for subsequent analysis, such as extracting features from the segmented regions and classifying breast density using a categorical scale.

In this section, we focus on the 2D breast density segmentation methods. Table 2.3 shows a summary of the existing approaches based on different techniques. Some methods are based on thresholding. Byng *et al.* (1994) proposed an interactive thresholding technique. Operators viewed mammographic images on a CRT display and selected a greylevel threshold to identify dense tissue within the breast. This interactive thresholding method has been widely applied in quantitative analysis of breast density. Since it is an interactive method, the operators must be trained and the process of setting a proper threshold can be time consuming. Moreover, as the threshold is subjectively selected, observer bias may be introduced. Instead of semi-automatic segmentation, some fully automatic methods have been developed where the threshold is determined automatically. Sivaramakrishna *et al.* (2001) developed a completely automatic method to segment dense regions using Kittler's optimal threshold (Kittler & Illingworth, 1986). Zhou *et al.* (2001) also used automatically determined thresholds to segment dense tissue from the breast region. Two threshold selection methods were employed: one was the Discriminant Analysis method (Otsu, 1975) where the optimal threshold is estimated using a discriminant criterion based on the intra-class variance and the inter-class variance, which is very appropriate for bimodal histograms; the other is the Maximum Entropy Principle based method (Wong & Sahoo, 1989) where the optimal threshold is determined by maximising a posteriori entropy, which can provide a better thresholding result for multimodal histograms. Chang *et al.* (2002) selected the median of all the pixel values within the breast region as the threshold to segment the breast region into two sub-regions: more dense and less dense. Torrent *et al.* (2008) described a multiple thresholding algorithm based on the excess entropy where the histogram was quantised by using the maximisation of the excess entropy. However, none of these automatic thresholding

methods is currently in widespread use (Yaffe, 2008). Note that these thresholding based methods only analyse the greylevel distribution of single pixels and apply a threshold to the image globally, without taking into account the aspect of spatial information of neighbouring pixels or local texture information.

Miller & Astley (1992) first investigated the adoption of texture analysis to discriminate glandular and fatty tissue in mammograms. They used granulometry and Laws' texture energy method (Laws, 1980) to classify breast tissue. Granulometric analysis was based on greylevel opening operations, and texture energy images were generated by convolving the original image with a number of small masks representing different texture features (e.g. edges, lines, spots or ripples). The obtained results indicated that texture analysis can provide a good basis for the automatic classification of breast tissue. The most significant error happened when the glandular disc had a relatively homogeneous appearance which was incorrectly classified as fatty tissue. After that, a number of texture based segmentation methods were presented. Zwiggelaar *et al.* (2003) combined texture features with expectation maximisation (EM) clustering (Dempster *et al.*, 1977) for segmenting mammographic images. A set-permutation-occurrence matrix was introduced to incorporate spatial information. The extracted texture features were then analysed to generate a statistical model. Finally, the EM clustering algorithm was performed to segment the breast region into six classes. Subsequently, Zwiggelaar & Denton (2006) developed a novel texture based segmentation method to segment mammograms into regions associated with four Wolfe patterns. This method was based on the principles that the ability of human observers to distinguish between textures is related to the contrast between the main structural elements and the repeating patterns formed by those specific structures. In this sense, they modelled local structural elements within specific greylevel bands to achieve the segmentation. Petroudi & Brady (2006) segmented mammographic images into different density regions using texture based statistical modelling. They applied textons in a hidden Markov random field (HMRF) framework. Textons were generated by filtering the mammographic images with a filter bank and clustering the filter responses. Subsequently, the texton feature vectors were analysed in the multi-vector Gaussian hidden Markov random fields to realise breast density segmentation. A closely related segmentation method was presented by Gong *et al.* (2006), which also combined statistical texture modelling with the HMRF algorithm. Instead of generating the textons from the filter responses of the filter bank, the textons were learnt by clustering image patch vectors extracted from a square neighbourhood around each pixel. With regard to statistical modelling of local texture, Zwiggelaar (2010) recently proposed a segmentation method based on local greylevel appearance (LGA) histograms. The local texture information was

modelled by transforming local greylevel appearance into a unique LGA number. The distribution of local greylevel appearances in histogram format constructed the basis for texture models. The LGA histograms generated from mammograms belonging to the same BIRADS class were combined to build four BIRADS models. These models were used to segment mammograms into four BIRADS density regions by labelling each pixel with one corresponding BIRADS class.

In addition, some approaches apply statistical analysis to obtain breast density segmentation. Heine & Velthuizen (2000) proposed a statistical methodology based on a chi-square probability analysis to discriminate fibroglandular and fatty tissue. It was assumed that there are two random events occurring in the input field of mammograms: one is fatty tissue with the smaller variance, and the other is non-fatty tissue with the larger variance. The variance of the image was computed over a small moving window. A global reference variance was determined to label each pixel as fatty or non-fatty. This method resulted in a region-wise continuous labelling of fatty and non-fatty tissue across the image. Ferrari *et al.* (2004c) proposed a method for the segmentation of the fibroglandular disc based on a statistical model of breast density. This breast density model was a mixture of Gaussians, estimated using the intensity distribution, representing different density categories. The EM algorithm and the minimum description length method (Rissanen, 1978) were used to determine the parameters of the density model and the number of tissue classes to be segmented. Adel *et al.* (2007) considered a mammographic image as a realisation of a random field and applied Markov random field (MRF) to segment regions of interest in the mammographic image. Bayesian theory and a maximum a posteriori method were used to estimate the parameters of the MRF model. Oliver *et al.* (2010) proposed a statistical method to segment the breast into fatty and dense tissue. Both tissue types were modelled based on statistical analysis of each pixel neighbourhood within the breast region. First, the same number of image patches were selected to represent fatty and dense tissue. Subsequently, two different statistical strategies, the Karhunen-Loeve transform and linear discriminant analysis were applied to build models for both tissue types. Finally, the built models were used to segment mammograms by classifying each pixel as fatty or dense tissue.

In contrast to the methods above, Bakic *et al.* (1998) developed a segmentation method by modelling breast tissue, compression effects, and mammogram formation. Fatty tissue was modelled by randomly distributed thin spherical shells, and fibroglandular tissue was modelled by connective tissue matrix containing small fatty compartments, ducts, and terminal ductal lobular units (Bakic *et al.*, 2002a,b). The generated breast model was used to select appropriate texture operators and spatial scales for segmenting fibroglandular tissue. Saha *et al.* (2001)

defined a fuzzy relation called fuzzy affinity to measure the connectivity between pixels. In this case, the dense region was seen as a fuzzy connectivity scene. An automatic connectedness threshold selection method was applied to segment the breast region into dense and fatty regions.

Some other methods are based on pixel-wise classification. Lao & Huo (2009) presented a method for the segmentation of dense tissue in mammograms based on unsupervised learning and multiple Levels of Detail. Firstly, an entropy maximisation based thresholding was applied to obtain a preliminary segmentation. Subsequently, fuzzy c-means clustering using the preliminary segmentation for the membership function intialisation was performed to generate a more accurate segmentation. Finally, a dense tissue feature was defined and computed for each pixel to obtain the final segmentation. Kallenberg *et al.* (2011) developed a pixel classification based segmentation method, which integrated different approaches in the literature to segment breast density in mammograms. A number of features including location, intensity, texture and global context were extracted. For breast density segmentation, an ensemble of five neural networks were trained using a separate training set that were already manually labelled. The sequential floating forward selection algorithm was applied to select an optimal subset from the whole feature space. On the other hand, some methods segment mammograms into four regions corresponding to four mammographic building blocks (i.e. nodular, linear, homogeneous, and radiolucent) as described by Tabár's tissue model. He *et al.* (2008) developed a texture based segmentation method using textons, which were generated by clustering the filter responses of a filter bank containing Gaussian, Laplacian of Gaussian and second-order derivatives of Gaussian filters at four scales. Texton selection was used to select discriminating textons between different tissue types and achieve optimised segmentation results. Subsequently, He *et al.* (2010) proposed a texture signature based segmentation method. Texture features were first extracted using a stack of annotated mammographic image patches. Tabár mammographic building blocks were modelled as texture signatures, and a model selection was performed to remove noise and outliers. Finally, the segmentation was achieved by labelling each pixel according to the generated models of Tabár mammographic building blocks. In addition, they developed a method based on geometric moments (He *et al.*, 2011). The procedure of this method was similar to the previous methods. Firstly, local image properties were extracted using the annotated image patches based on local moments. Tabár mammographic building blocks were then modelled by clustering the obtained feature vectors. Finally, the generated models were used to segment mammographic images. By contrast, He *et al.* (2012) recently proposed a novel mammographic density segmentation approach, which segmented the breast into dense, semi-dense,

Table 2.3: A summary of breast density segmentation methods, indicating authors, years of publication, techniques, experimental data for test, and segmentation performance.

Author & Year	Thresholding	Clustering	Statistical	Texture	Modelling	Classification	Test Data	Segmentation Performance
Miller & Astley (1992)		√		√		√	40	80% correct classification rate*
Byng *et al.* (1994)	√						60	Spearman coefficient $\rho > 0.91^{\dagger}$
Bakic *et al.* (1998)				√	√		unknown	capable
Heine & Velthuizen (2000)			√				50	most were verified by experts
Sivaramakrishna *et al.* (2001)	√						32	Spearman coefficient 0.92 - $0.95^{\dagger}$
Zhou *et al.* (2001)	√						260	correlation 0.94 CC; 0.91 MLO†
Saha *et al.* (2001)	√						174	100% visually acceptable
Chang *et al.* (2002)	√						200	PCC $r > 0.8^{\dagger}$
Zwiggelaar *et al.* (2003)		√	√	√	√		MIAS	86% agreement†
Ferrari *et al.* (2004c)		√	√		√		84 MIAS	64.3%, 16.7%, 10.7%, 4.7%, 3.6%§
Zwiggelaar & Denton (2006)			√	√	√		60	72% agreement (Wolfe)†
Petroudi & Brady (2006)		√	√	√	√		32	28/3/1; 96.7%/93.8%/92.8%$^{\diamond}$
Gong *et al.* (2006)		√	√	√	√		43	97%/95.7%/91.9%/97.2%$^{\heartsuit}$
Adel *et al.* (2007)			√		√		50 MIAS	68% rated as good
Torrent *et al.* (2008)	√	√	√				250	$Az = 0.818$ MLO; $Az = 0.803$ CC
He *et al.* (2008)		√		√	√	√	full MIAS	realistic with Tabár's model
Lao & Huo (2009)	√	√				√	220 cases	PCC $r = 0.78$ MLO vs CC
Zwiggelaar (2010)		√		√	√	√	full MIAS	64% agreement (BIRADS)
Oliver *et al.* (2010)			√		√	√	322 MIAS 60 Trueta	feasible/robust Accuracy up to 0.916 ± 0.038*
He *et al.* (2010)		√		√	√	√	full MIAS	better match Tabár's model
Kallenberg *et al.* (2011)				√		√	750	PCC $r = 0.911, 0.895^{\ddagger}$
He *et al.* (2011)		√		√	√	√	full MIAS	42%, 58%, 69%, 96%$^{\triangle}$
He *et al.* (2012)		√	√		√	√	full MIAS	CA 85% Tabár; 78% BIRADS
Keller *et al.* (2011)		√	√			√	160	PCC $r = 0.83^{\dagger}$
Keller *et al.* (2012)		√	√			√	81 women	PCC r up to 0.85 per image† PCC r up to 0.89 per woman†

PCC: Pearson correlation coefficient; Az: the area under the ROC curve; CA: classification accuracy.

* pixel-based classification.

† with respect to (average) subjective classification, assessment, segmentation, or measurement of mammographic density by expert(s).

§ 64.3% were rated as excellent, 16.7% good, 10.7% average, 4.7% poor, 3.6% complete failure.

◇ 28 very satisfactory, 3 satisfactory, 1 good (image-based visual rating); segmentation accuracy: $96.7 \pm 3.4\%$ for dense tissue, $93.8 \pm 2.9\%$ for fatty tissue, and $92.8 \pm 4.2\%$ for breast edge (pixel-based quantitative evaluation).

♡ segmentation accuracy: $97 \pm 2.7\%$ for dense tissue, $95.7 \pm 2\%$ for dense tissue with structures, $91.9 \pm 2\%$ for fatty tissue, and $97.2 \pm 3.6\%$ for fatty edge (pixel-based quantitative evaluation).

‡ Pearson correlation coefficients were 0.911 and 0.895 for percent density and dense area with the user-assisted threshold method.

△ clinically acceptable rate was 42% for Pattern I, 58% for II/III, 69% for IV, and 96% for V (image-based visual rating).

and non-dense tissue instead of Tabár's four building blocks. Greylevel histogram based features were derived from multi-resolution local windows and used to build models of Tabár mammographic building blocks. A Bayesian classifier was trained based on a novel binary model matching pattern to achieve mammographic segmentation. The tissue class of every pixel was determined by calculating the probability of it belonging to each class.

Keller *et al.* (2011) applied a multiclass fuzzy c-means clustering algorithm to segment the breast into regions of similar greylevels, in which the number of clusters for a specific mammogram was automatically determined based on the number of modes contained in the greylevel histogram. The final segmentation of the dense versus fatty tissue was achieved by merging clusters into two classes using a linear discriminant analysis agglomeration classifier, where histogram statistics, image acquisition parameters, and patient characteristics were involved. Subsequently, Keller *et al.* (2012) modified their previous work and employed a support vector machine classifier for dense tissue segmentation, which was trained to identify clusters corresponding to dense tissue based on a combined feature set.

With regard to breast density segmentation, we attempt to develop two methods mainly based on intensity information in this thesis. This follows the fact indicated by the studies above that intensity levels in mammograms directly reflect breast tissue densities. Compared to the existing two major types of breast density segmentation methods, the first method is regarded as multiclass breast density segmentation, which partitions pixels in the breast region into multiple groups corresponding to different densities. The second method is intended to distinguish dense tissue from non-dense (fatty) tissue. To achieve multiclass breast density segmentation, using clustering algorithms tends to be very straightforward where pixels with similar intensity levels are grouped into clusters and as such the entire breast region can be segmented into sub-regions with different densities. Hence we apply the fuzzy c-means algorithm for multiclass breast density segmentation. To eliminate the sensitivity to noise and intensity inhomogeneities, we develop a modified fuzzy c-means algorithm by incorporating spatial information, which will be presented in Section 3.3. To detect dense tissue regions, we focus on regions with higher intensity levels and propose a method based on topographic representation of the breast region, which will be presented in Section 3.4.

2.2 Estimation of Mammographic Density

As mentioned in the above section, quantitative estimation of breast density is beneficial to mammographic risk assessment. In this section, methods for estimating mammographic density are reviewed. A summary can be found in Table 2.4.

Most of these are based on the area of dense tissue regions in the breast. Byng *et al.* (1994, 1998) proposed percent density (PD) for measuring mammographic density, which is the proportion of dense tissue within the breast, obtained by dividing the area of dense tissue by the area of the entire breast. The boundaries of dense tissue regions and the breast were identified using interactive thresholding. Similarly, Ursin *et al.* (1998) used the percentage of dense tissue in the breast as a measure of mammographic density where dense tissue regions were also determined by interactive thresholding. Changes in mammographic density over 12 months were monitored based on the measure. Heine *et al.* (2008) also used an area based measure, produced by an automated method for mammographic density detection (Heine & Velthuizen, 2000). The correlation between the two measures resulted from the user-assisted thresholding (Byng *et al.*, 1994, 1998) and the automated breast density detection method was investigated. Martin *et al.* (2006) used percentage density to measure mammographic density, which is similar to PD, calculated by dividing the number of pixels in the dense region by the total number of pixels in the entire breast region. They compared the resulting percentage density using a computer-aided method (Zhou *et al.*, 2001) and that estimated by radiologists. The BIRADS density categories were used to evaluate the reproducibility of breast density quantification. Lu *et al.* (2007) modified the interactive thresholding based method of Byng *et al.* (1994, 1998) for measuring breast density in full-field digital mammograms. Mathematical models were constructed by incorporating image pixel intensity statistics and instrument parameters for image acquisition, in order to objectively predict threshold values and remove the observer based segmentation step. Keller *et al.* (2012) developed a fully automatic algorithm for estimating breast PD in both raw and post-processed digital mammography images. Strong correlation was indicated between the algorithm-estimated breast PD and the radiologist-provided breast PD for both raw and processed digital mammograms. In addition to digital mammograms, central digital breast tomosynthesis (DBT) projection images were analysed. Bakic *et al.* (2009) evaluated inter/intra-reader agreement in breast density estimation based on PD which was estimated by using Cumulus software (Byng *et al.*, 1994, 1998). In contrast to the studies above, Stone *et al.* (2009) investigated two measures of mammographic density: one was dense area defined as the number of pixels in the dense region; the other was percent dense area calculated as a ratio of dense area and the total area of the breast, i.e. proportion/percentage of dense tissue. Multivariate linear regression was employed to evaluate associations between the two measures of mammographic density and breast cancer risk factors. Dense area was found to be as reliable as that of percent dense area. Moreover, dense area was not associated with other risk factors, making it a simpler biomarker

for risk prediction modelling. Saha *et al.* (2001) defined different measures for characterising mammographic density, which were computed from the segmented dense tissue regions, including the sum of intensity values of pixels in the dense region, the sum of intensity values of pixels in the fatty region, the ratio between them, the area of the dense region, the ratio between the areas of the dense and fatty regions, the ratio between the areas of the dense region and the breast, etc. In addition, Raundahl *et al.* (2006a,b) proposed to measure density changes using tissue structural information orthogonal to intensity based methods, which was based on an assumption that density changes can be perceived as a structural matter that may be characterised without considering the brightness of the image. Features derived from eigenvalues of the Hessian matrix were used to classify pixels into four classes and as such the breast was segmented into four sub-regions corresponding to different anatomical structures. A linear combination of the relative areas of the four classes was exploited to determine the density score of the breast. Subsequently, they applied this method to quantify changes in mammographic appearance under different physiological processes, such as aging and hormone replacement therapy (Raundahl *et al.*, 2008).

On the other hand, some studies have reported that measuring breast density based on the projected area of dense tissue in 2D mammography is not ideal, as different compression and projection angles can induce variability. Hence the 3D approaches have been developed to estimate breast density based on the volume of dense tissue. Highnam & Brady (1999) developed a mathematical model of mammographic imaging to model the process of generating a mammographic image. A quantitative measure of breast tissue at each pixel location in the image can be constructed automatically on the basis of this mammography model. This measure was referred to as h_{int}, representing the thickness of "interesting" tissue (non-fatty tissue). As such, a representation of the non-fatty tissue can be derived from a single mammographic image by calculating the h_{int} measure over the image, from which the volume of the non-fatty tissue and breast density can be quantitatively estimated. Marias *et al.* (2004) applied the h_{int} representation which they called the Standard Mammogram Form (SMF) representation to characterise breast density changes due to hormone replacement therapy. Blot & Zwiggelaar (2005) proposed an approach to the volumetric estimation of breast density based on the h_{int} representation. Two measures were evaluated: the first was the volume of the non-fatty tissue defined as the sum of h_{int} values over the breast; and the second was the ratio of the volumes of the non-fatty tissue and the breast. Subsequently, Highnam *et al.* (2006) revised their previous theoretical analysis of SMF with regard to errors in the calibration parameters. It was shown that by estimating breast thickness the errors in the calibration data can be ef-

fectively compensated and SMF can be regarded as a promising way of analysing screen-film mammograms with little or no calibration data.

Instead of using a physics based model to estimate the thickness of the non-fatty tissue (Highnam & Brady, 1999), Pawluczyk *et al.* (2003) developed a technique for breast density quantification fully based on calibration data. A plastic calibration device composed of a range of tissue equivalent materials was first employed to estimate operating parameters of the imaging system. Subsequently, variations in exposure factors and film processing characteristics were corrected through images using the aluminum step wedge. Finally, the optical density information at each pixel in a digitised film mammogram was used to estimate the exposure transmitted by the breast and the tissue composition that caused this transmission, which enabled an estimation of the volume of dense tissue, the total volume of the breast, and their ratio, i.e. volumetric breast density. This technique was further evaluated by Boyd *et al.* (2009) using sixteen calibrated mammography machines in seven locations in Toronto. The tissue composition (the proportion of fatty and dense tissue) at each pixel in the screen-film mammogram was estimated according to the measured image signal and a measurement of breast thickness was incorporated to calculate the volume of dense tissue.

In contrast to estimating volumetric breast density in digitised mammograms, van Engeland *et al.* (2006) proposed a method for estimating the volume of dense tissue in full-field digital mammograms, where the thickness of dense tissue at a pixel was determined by using a physical model of image acquisition. Specifically, a polychromatic image model was designed, which incorporated the photon energy spectrum for different anode target/filter materials and the energy dependency of linear attenuation coefficients of breast tissue. Effective linear attenuation coefficients of fatty and dense tissue were computed from empirical data as a function of tube voltage, anode material, filtration, and compressed breast thickness, which were subsequently used in a monochromatic computation of the volume of dense tissue. Hartman *et al.* (2008) improved the algorithm of Highnam & Brady (1999) and proposed a new algorithm for the volumetric estimation of breast tissue composition for digital mammograms, which can be regarded as a digital version of Highnam *et al.* (2006). This algorithm took advantage of the high image quality and the additional information of digital mammograms and incorporated the recorded breast thicknesses into the image based calibration.

Instead of using mammograms for estimating volumetric breast density, Wei *et al.* (2004) developed a semi-automatic method for estimating the percent volume of dense tissue using breast MR images. The breast boundary was first detected automatically on each slice. Manual interactive thresholding was then used to segment dense tissue from each slice. The volume of dense tissue was calculated

Table 2.4: A summary of breast density estimation methods, indicating authors, years of publication, breast density measures, test data, and performance.

Author & Year	Area	Volume	Structure	Test Data	Indication of the Performance
Byng *et al.* (1994)	√			60	Spearman coefficient $\rho > 0.91^{\dagger}$
Byng *et al.* (1998)	√			60	intra-class correlation > 0.9
Ursin *et al.* (1998)	√			19 women	Pearson correlation coefficient $r > 0.85^{\dagger}$
Highnam & Brady (1999)		√		N/A	N/A
Saha *et al.* (2001)	√			174	two-projection correlation > 0.9
Pawluczyk *et al.* (2003)		√		phantom	difference $< 5\%$ with actual value
Marias *et al.* (2004)		√		59 pairs	agreement with clinician up to 86%
Wei *et al.* (2004)		√		67 cases	$r = 0.91$ (MR/CC); $r = 0.89$ (MR/MLO)
Blot & Zwiggelaar (2005)		√		N/A	N/A
Martin *et al.* (2006)	√			260	Pearson correlation coefficient $r = 0.89^{\dagger}$
Highnam *et al.* (2006)		√		4028	up to 99.3% excellent estimation
Raundahl *et al.* (2006a)			√	50 patients	separate HRT and placebo patients
Raundahl *et al.* (2006b)			√	artificial	differentiate HRT and placebo patients
van Engeland *et al.* (2006)		√		22 patients	$r = 0.94, 0.97$; relative error $13.6\%^{*}$
Lu *et al.* (2007)	√			85	intra-class correlation up to 0.94
Heine *et al.* (2008)	√			703 patients	$r = 0.70$ with user-assisted thresholding
Hartman *et al.* (2008)		√		1277	$r = 0.92$ for 275 pairs of left/right*
Raundahl *et al.* (2008)			√	unknown	comparable to the best method
Bakic *et al.* (2009)	√			39 women	inter-reader $\rho = 0.85 \pm 0.05$ on DBT intra-reader $\rho = 0.88 \pm 0.05$ on DBT
Stone *et al.* (2009)	√			799	$r = 0.75; r = 0.85^{\S}$
Boyd *et al.* (2009)		√		364/656	odds ratio $1.98/1.68^{\diamondsuit}$
Keller *et al.* (2012)	√			81 women	r up to 0.85 per image† r up to 0.89 per woman†

† with respect to manual estimation of mammographic density by expert(s).
* the correlation between MRI and mammography volumes was 0.94 per image and 0.97 per patient; the relative error was evaluated with respect to MRI.
* 448 images correlated well with a reader's visual assessment.
§ $r = 0.75$ between percent dense area and dense area; $r = 0.85$ between the computer-assisted and the visual assessment of percent dense area.
$^{\diamondsuit}$ 364 cases and 656 controls; the odds ratio between the 5^{th} quintile and the 1^{st} quintile was 1.98 for the volume measure and 1.86 for the area measure.

by integrating the segmented dense area over all the slices containing the breast. Moreover, they investigated the correlation between the percent volume of dense tissue in breast MR images and the percent dense area in corresponding mammograms. A high correlation was found between the percent volume and the percent area of dense tissue in the breast.

In this thesis, we investigate the role of quantitative estimation of breast density in mammographic risk assessment. Two quantitative measures, *density* and $BLOB_{norm}$ (see Sections 3.4.3 and 5.3.3), are defined based on the segmentation of dense tissue regions and the representation of dense tissue patterns, respectively.

2.3 Characterisation of Mammographic Parenchymal Patterns

Mammographic parenchymal patterns have been found to be an important indicator of the risk of developing breast cancer. A strong relation between mammographic parenchymal patterns and breast cancer risk has been described by many studies. A wide variety of methods for computerised analysis of mammographic parenchymal patterns have been developed in order to obtain automatic, objective and quantitative mammographic risk assessment. In this section, methods for the characterisation of mammographic parenchymal patterns are reviewed, where different types of features were extracted and used to classify parenchymal patterns into different risk levels. Table 2.5 provides a summary of these methods, indicating the techniques used for feature extraction.

Caldwell *et al.* (1990) characterised mammographic parenchymal patterns using fractal dimension to distinguish between cases associated with a high risk of breast cancer and those associated with a low risk of breast cancer. A quantitative measure of parenchymal patterns in mammograms may be defined based on this fractal based parameter to assess changes in parenchymal patterns for a long-term study of patients with a high breast cancer risk.

Tahoces *et al.* (1995) used three groups of features to characterise mammographic breast parenchymal patterns. Small regular square ROIs and large irregular ROIs were manually selected for feature extraction. The first and second groups of features were based on the Fourier transform and local-contrast analysis, which were derived from the regular ROIs, in order to show the amount of dense tissue within the ROI and discriminate different types of breast tissue architecture. The third group was based on the greylevel distribution in the irregular ROIs, which was applied to distinguish dense from fatty breasts.

Byng *et al.* (1996a) described mammographic parenchymal patterns based on variations in both greylevel and texture. Two features (fractal dimension and skewness) were extracted to characterise these variations. The fractal dimension was used to describe texture information and the skewness was used to represent the asymmetry of the greylevel histogram.

Karssemeijer (1998) divided the breast area into regions in which the distance of pixels to the breast boundary was approximately equal. To segment these regions automatically, a distance transform was used. The distance of a pixel to the skin line was determined by eroding the breast mask repeatedly using a circular structuring element until this pixel was removed from the mask, and the number of erosions represented its distance to the skin line. Greylevel histograms were computed for these regions and two features (standard deviation and skewness) were

calculated from these histograms. In addition, the probability density functions of the pixel values were estimated in the pectoral and breast regions, from which two features were extracted to measure the difference of the two distributions.

Huo *et al.* (2000) extracted 14 features from the central breast region to characterise mammographic parenchymal patterns. This feature set comprised four groups: features based on the absolute greylevel values, features based on greylevel histogram analysis (balance and skewness), features based on the spatial relationship among greylevels within the ROI (contrast, coarseness, and features derived from co-occurrence matrices), and features based on Fourier analysis. These features were used to discriminate between mammographic patterns observed from women at a low breast cancer risk and those observed from women having a very high risk of developing breast cancer. Moreover, four features were selected and related to the risk as estimated from existing clinical models to investigate the capability of the extracted features in predicting breast cancer risk. After that, Li *et al.* (2005) improved their work by adding two groups of new features to the feature set, which were features based on fractal analysis and features based on edge frequency, respectively. Similarly, the extracted features were employed to distinguish between mammographic parenchymal patterns of women at low breast cancer risk and those of women at high risk. Recently, they extended the evaluation using a large clinical dataset of full-field digital mammograms (Li *et al.*, 2012). It was demonstrated again that women at high risk of developing breast cancer tend to have dense breasts with coarse and low contrast texture patterns.

Blot & Zwiggelaar (2001) proposed an approach to classifying mammographic parenchymal patterns using only the background texture information in mammograms. The background texture and image structures were separated based on the difference between local co-occurrence matrix and mean co-occurrence matrix. The local co-occurrence matrix was calculated from a local square region, while the mean co-occurrence matrix was generated by averaging eight co-occurrence matrices computed from eight neighbouring square regions around the local region. This was based on an assumption that there is a statistical difference between the greylevel co-occurrence matrices for image regions containing image structures and those for background regions only containing background texture. Finally, the features defined by Karssemeijer (1998) were extracted from the reconstructed background texture for classification.

Marias *et al.* (2002) investigated the role of quantitative measures of mammographic density in characterising the breast parenchyma. The h_{int} representation (Highnam & Brady, 1999) of mammograms was used for calculating two measures to classify mammographic patterns. The first one was the percent volume of dense tissue, and the second one was the correlation measure derived from

the co-occurrence matrix of the h_{int} representation.

Zwiggelaar *et al.* (2004) concentrated on linear structures and presented four methods for extracting linear structures from mammograms, including line operator, orientated bins, Gaussian derivatives, and ridge detector. All these methods can estimate line-strength, orientation and scale at each pixel in the mammogram. Subsequently, Hadley *et al.* (2006) related the density of linear structures in mammograms to mammographic risk assessment. A measure of line-strength was computed for each pixel using the line operator method of Zwiggelaar *et al.* (2004), and a threshold was set on the line-strength measure to remove the background texture. The proportion of pixels with line-strength values above the threshold was computed and used for distinguishing different risk categories.

Petroudi *et al.* (2003) modelled mammographic parenchymal patterns as a statistical distribution of textons which were obtained by clustering filter responses of a filter bank. A set of training mammograms were selected covering different classes of parenchymal patterns. The MR8 filter bank was used to filter the segmented breast area and the k-means algorithm was employed to compute the cluster centres, representing the textons of the mammographic appearance. A texton histogram was computed for each training mammogram by assigning a texton to each breast tissue pixel and counting the occurrence of each texton over the mammogram. The texton histograms belonging to the same class constituted the models of parenchymal patterns for that class.

Kinoshita *et al.* (2007) used a wide range of features to characterise breast density patterns as well as the anatomical size and shape of the breast in mammograms, including shape features, Haralick's texture features, Hu's moment based features, statistics of greylevel histograms, granulometric features, and Radon domain features. The shape features were extracted from the contour of the breast region; the texture features were derived from the spatial greylevel dependence matrix of the breast region; the moment based features and the statistics of greylevel histograms were computed from two sub-regions of the breast segmented by using the nipple as reference; the granulometric features were extracted from fibroglandular tissue; and the Radon domain features were derived from the Radon transform functions of fibroglandular tissue in the two sub-regions.

Castella *et al.* (2007) made use of a set of statistical features to describe four manually selected ROIs for each mammogram. These included features derived from greylevel histograms (standard deviation, skewness, kurtosis, and balance), features extracted from greylevel co-occurrence matrices (energy, entropy, cmax, contrast, and homogeneity), features based on the primitive matrix (short/long primitive emphasis, greylevel uniformity, and primitive length uniformity), fractal dimension, and features computed from the neighbourhood greytone difference

matrix (coarseness, contrast, complexity, and strength).

Li *et al.* (2008) applied power spectral analysis to characterise mammographic parenchymal patterns. An ROI was first manually selected from the central breast region behind the nipple. The discrete Fourier transform was then performed on the ROI and the corresponding power spectrum was calculated. The exponent β of the power law spectrum was estimated and used as a feature for breast cancer risk assessment. It was shown that mammographic parenchymal patterns of high-risk women indicated a larger β value than those of low-risk women.

Manduca *et al.* (2009) evaluated the association of five families of texture features with breast cancer risk, which were derived from Markovian co-occurrence matrices, run-length analysis, Laws features, wavelet decomposition, and Fourier analysis, respectively. Markovian features were used to analyse how frequently pixels of given greylevel values appear in an image with a specific distance; run-length features were used to examine runs of similar greylevel values in an image; Laws features were computed from a set of filters to capture different types of structures in an image; wavelet features were employed to enhance broader diagonal, vertical, and horizontal textures; and Fourier features were applied to describe the coarseness of textures.

He *et al.* (2010) proposed a texture signature based methodology to model mammographic building blocks as described in Tabár's model. The texture signature is composed of three 2D cumulative histograms, generated from local circular windows. The first 2D histogram describes the greylevel distribution of pixels at different radial distances from the central pixel within local windows; the second 2D histogram represents the greylevel distribution of pixels at different angles with respect to the central pixel within local windows, and the third 2D histogram is a specific co-occurrence matrix with $d = 1$ and $\theta = 0°$, representing the magnitude of the greylevel variance between two adjacent pixels.

Nie *et al.* (2010) analysed breast parenchymal patterns using three-dimensional breast MRI. The fibroglandular tissue was segmented from all slices and four morphological features (circularity, convexity, irregularity, and compactness) were computed based on the segmented fibroglandular tissue, which were subsequently used to classify cases into two breast parenchymal patterns: intermingled pattern with intermixed fatty and fibroglandular tissue and central pattern with confined fibroglandular tissue inside surrounded by fatty tissue outside.

Wei *et al.* (2011) designed a quantitative measure of mammographic parenchymal patterns to analyse the texture patterns of fibroglandular tissue. An ROI was first located by using the nipple as reference, covering the main area of the breast parenchyma. Ten run-length statistics and five region-size statistics features were then extracted from the ROI. Stepwise linear discriminant analysis

Table 2.5: A summary of methods for characterising mammographic parenchymal patterns, indicating authors, years of publication, the techniques used for feature extraction, experimental data for test, and indication of the performance.

Author & Year	Fractal	Greylevel Statistics	Co-occurrence Matrix	Domain Transform	Texture Modelling	Structure/Morphology	Test Data	Indication of the Performance
Caldwell *et al.* (1990)	√						70	84.3% agreement with radiologists
Tahoces *et al.* (1995)		√		√			117	$\kappa = 0.52/0.22/0.52/0.77$ (Wolfe)
Byng *et al.* (1996a)	√	√					60	$r = -0.88$ skewness; $r = -0.76$ fractal†
Karssemeijer (1998)		√					615	67% agreement with a radiologist
Huo *et al.* (2000)		√	√	√			356 women	Az up to 0.91*
Blot & Zwiggelaar (2001)		√	√				265 MIAS	63% success rate
Marias *et al.* (2002)			√				132 MIAS	classification accuracy up to 91%
Petroudi *et al.* (2003)					√		132	76% classification accuracy
Zwiggelaar *et al.* (2004)						√	unknown	$Az = 0.943$; $Az = 0.746^{\star}$
Li *et al.* (2005)	√	√	√	√			172 women	Az up to 0.86 (low/high risk)
Hadley *et al.* (2006)						√	321 MIAS	inter-class p-values < 0.05
Kinoshita *et al.* (2007)		√		√			1080	precision rate 79% − 83%
Castella *et al.* (2007)	√	√	√				352	κ up to 0.83 ± 0.08
Li *et al.* (2008)				√			172	$Az = 0.90$ (low/high risk)
Manduca *et al.* (2009)		√	√	√	√		387 women	$Az = 0.58 - 0.60$ predict cancer
He *et al.* (2010)					√		full MIAS	better match Tabár's model
Nie *et al.* (2010)						√	230 patients	Az up to 0.94 (two distinct patterns)
Wei *et al.* (2011)		√					165 women	$Az = 0.74 \pm 0.05$ (case/control)
Wei *et al.* (2012)		√		√			123 subjects	$r = 0.02 - 0.67$ FFDM/DBT
Li *et al.* (2012)	√	√	√	√			456 women	$Az = 0.82$; $Az = 0.73^{\ddagger}$

Greylevel statistics include features extracted from greylevel histograms (i.e. standard deviation, skewness, kurtosis, and balance), features based on the spatial relationship among greylevels (i.e. contrast, coarseness, run-length statistics, and region-size statistics), and features based on the primitive matrix (i.e. short/long primitive emphasis, greylevel uniformity, and primitive length uniformity), etc. Co-occurrence Matrix represents features derived from co-occurrence matrices (i.e. energy, entropy, contrast, and homogeneity). Domain transform indicates features extracted by using the Fourier or Radon transform.

† with respect to radiologists' subjective classifications of mammographic parenchymal patterns.

* the area under the ROC curve (Az) was up to 0.91 for the classification of low-risk women and BRCA1/BRCA2 mutation carriers.

* $Az = 0.943$ for the pixel-level detection of linear structures; $Az = 0.746$ for the classification of linear structures into target class (duct or spicule) and non-target class.

‡ $Az = 0.82$ for the classification of BRCA1/2 gene mutation carriers and low-risk women; $Az = 0.73$ for the discrimination between unilateral cancer and low-risk women.

was used to search for the best combination of features, and a linear discriminant function was defined using the subset of features, which was applied to generate the measure of mammographic parenchymal patterns for each mammogram. Furthermore, Wei *et al.* (2012) investigated the correlation between the measures of breast parenchymal patterns extracted from full-field digital mammography and those extracted from digital breast tomosynthesis, including run-length statistics, region-size statistics, and power spectrum.

As shown in Table 2.5, various features have been used for characterising mammographic parenchymal patterns. Most features are based on texture information in mammograms, due to the fact that parenchymal patterns are interpreted as textures in computer vision. Therefore, we also propose modelling mammographic parenchymal patterns using texture features. A brief review of texture analysis approaches will be presented in Section 4.1 where various methods for texture feature extraction are discussed, covering four categories: statistical methods, structural methods, signal processing methods, and model based methods. Then we proceed to focus on local feature based texture representations (see Section 4.2), and investigate different types of local features for modelling mammographic tissue appearance (see Section 4.3). In fact, we intend to generate a dictionary of local features to describe local tissue appearance. As such, the overall appearance of mammographic tissue can be modelled with occurrence histograms by statistical analysis of local features according to the dictionary. In addition, we plan to explore a new idea for the representation of mammographic parenchymal patterns instead of using texture features. Motivated by Tabár's model of parenchymal patterns (Tabár *et al.*, 2005) and mammogram synthesis work (Bakic *et al.*, 2002a), we propose a new method for representing parenchymal patterns by focusing on blob-like dense tissue patterns in mammograms. A multiscale blob based representation of mammographic parenchymal patterns will be presented in Section 5.3.

2.4 Breast Density Classification

In the section above, we have reviewed a number of methods for the characterisation of mammographic parenchymal patterns, which represent parenchymal patterns based on a variety of features extracted from mammographic images. In this section, we mainly review the approaches to modelling breast tissue in mammograms and classifying breast density into different density classes. A summary of these approaches is shown in Table 2.6. The related publications among them will be used for a comprehensive comparison with our work in Chapter 8.

Bovis & Singh (2002) employed four groups of texture features, a series of statistical features, a circularity shape feature, and fractal dimension to characterise the

underlying texture contained within breast tissue in mammograms. Eleven artificial neural network (ANN) classifiers were trained using different input datasets to provide an estimate of the posterior probabilities for the four BIRADS classes. The classifier outputs were subsequently combined using six different combination rules. A comparison was made between the results achieved by using the combination of multiple classifiers and those for the best single classifier.

Petroudi *et al.* (2003) built a set of models for each BIRADS class in the form of texton histograms. For a test mammogram, a texton histogram was generated and compared with all the built models. The class label was determined according to the nearest model using the χ^2 distribution comparison.

Oliver *et al.* (2005a,b, 2006, 2008) developed a series of methods for breast density classification. Oliver *et al.* (2005b) used the k-means algorithm to group pixels in the breast into two classes (fatty and dense tissue). Morphological features and texture features based on co-occurrence matrices were extracted from both tissue classes for the subsequent breast density classification. The k-Nearest Neighbours (kNN) classifier and the ID3 decision tree classifier (Duda *et al.*, 2001b) were used to classify breast density as dense, fatty, or glandular. In the same year Oliver *et al.* (2005a) proposed a method for classifying breast density into the four BIRADS classes where the same features and classifiers were used. While for segmenting the breast into fatty and dense tissue, the fuzzy c-means algorithm was employed rather than using the k-means algorithm. Oliver *et al.* (2006) subsequently compared five different strategies for feature extraction according to the regions from which the features were extracted. The first strategy extracted features from the whole breast region, while the remaining four strategies extracted features from the sub-regions within the breast area, obtained by using the distance of each pixel to the skin line (Karssemeijer, 1998), fuzzy c-means, fractal analysis, and statistical analysis, respectively. Similarly, texture features were derived from co-occurrence matrices and a set of morphological features were calculated, including the relative area and the four first histogram moments. A Bayesian classifier combining the kNN and the C4.5 decision tree algorithms was used for classifying breast density into the four BIRADS categories. After that, Oliver *et al.* (2008) performed a comprehensive evaluation of their previous method using two public and widely known databases. The individual performance of the kNN classifier, the C4.5 decision tree classifier, and the combination of the two classifiers using the Bayesian approach was tested. In addition, the sequential forward selection algorithm (Kittler, 1986) was included into the kNN classification to select the most discriminant features and improve the classification performance. In addition, Mac Parthaláin *et al.* (2010) applied rough and fuzzy-rough set approaches to analyse the extracted feature data of Oliver *et al.* (2008). A number of clas-

sifiers including a fuzzy nearest neighbours (FNN) classifier (Keller *et al.*, 1985), a fuzzy-rough nearest neighbours (FRNN) classifier (Jensen & Cornelis, 2011), a fuzzy-rough ownership function based classifier (FRNN-O) (Sarkar, 2007), and a nearest neighbour classifier based on the vaguely quantified rough set model (VQNN) (Cornelis *et al.*, 2007) were employed for the BIRADS density classification. The impact of dimensionality reduction on classification accuracy was also examined by using two feature selection methods, fuzzy-rough feature selection (FRFS) (Jensen & Shen, 2009) and distance metric-assisted tolerance rough set feature selection (DMTRS) (Mac Parthaláin & Shen, 2009), respectively.

Zwiggelaar *et al.* (2005) built statistical models of different density classes on the basis of greylevel histograms of mammograms. First, the normalised greylevel histograms (64 bins) belonging to the same density class were grouped, which formed a model of this density class. Principal component analysis was subsequently applied to obtain the principal components of each model. The first few principal components of each model constructed a feature space for breast density classification. The kNN classifier was used to classify mammograms into three density classes (fatty, glandular, and dense).

Muhimmah & Zwiggelaar (2006) used a multi-resolution histogram technique instead of single-scale greylevel histograms. A five-level Gaussian pyramid was first built for each mammogram and greylevel histograms of the breast region were computed for all levels. Subsequently, cumulative histograms were generated and difference histograms between adjacent levels were computed. Finally, the resulting difference histograms were concatenated to form the multi-histogram feature space. A multiclass directed acyclic graph support vector machine (DAG-SVM) classifier was employed for breast density classification using three categories.

In contrast to most methods, Castella *et al.* (2007) assessed the density category of multiple ROIs in each mammogram using a set of statistical features extracted from these regions. Three different classifiers were employed for breast density categorisation of the selected ROIs, which were linear discriminant analysis (LDA), a classic Bayesian classifier based on Mahalanobis distance, and a naive Bayesian classifier, respectively. The density category for each mammogram was finally determined by averaging the individual ROI classification in the mammogram.

Boehm *et al.* (2008) classified breast parenchymal density into three categories: fibrosis, involution atrophy, and normal. Features were extracted from retromammilar image sections in which the density pattern was supposed to be more or less homogeneous. Two types of features were evaluated, based on intensity information and topologic analysis, respectively. For intensity based features, the 20^{th} percentile, median, and mean values were calculated from greylevel histograms. For topologic analysis, Minkowski functionals (Thompson, 1996) were used from

Table 2.6: A summary of breast density classification methods, indicating authors, years of publication, extracted features, classifiers, breast density classification schemes, experimental data for evaluation, and classification accuracy (CA).

Author & Year	Feature	Classifier	Scheme	Test Data	CA
Bovis & Singh (2002)*	statistical+shape+fractal dimension	ANN	BIRADS	377 DDSM	up to 71.4%
Petroudi *et al.* (2003)*	texton histogram	1NN	BIRADS	88	76%
Oliver *et al.* (2005b)	morphological+texture	kNN, ID3	Three Categories[1]	90 MIAS	up to 70%
Zwiggelaar *et al.* (2005)	greylevel histogram	PCA+kNN	Three Categories[1]	312 MIAS	71.5%
Oliver *et al.* (2005a)*	morphological+texture	kNN, ID3	BIRADS	300 DDSM	up to 47%
Oliver *et al.* (2006)*	morphological+texture	Bayesian	BIRADS	322 MIAS	up to 82%
Muhimmah & Zwiggelaar (2006)	multi-resolution greylevel histogram	DAG-SVM	Three Categories[1]	321 MIAS	77.57%
Castella *et al.* (2007)*	statistical	LDA, Bayesian	BIRADS	352	up to 80%
Oliver *et al.* (2008)*	morphological+texture	SFS+kNN, C4.5, Bayesian	BIRADS	322 MIAS 831 DDSM	up to 86% up to 77%
Boehm *et al.* (2008)	intensity+topologic	Discriminant Analysis	Three Categories[2]	100	up to 89%
Subashini *et al.* (2010)*	statistical	SVM	Three Categories[1]	43	up to 95.44%
He *et al.* (2010)*	Tabár's model based segmentation	kNN	BIRADS, Tabár	320 MIAS	75%; 78%
Mac Parthaláin *et al.* (2010)*	morphological+texture	FNN, FRNN, FRNN-O, VQNN	BIRADS	322 MIAS 831 DDSM	up to 91.4% up to 89.24%
Bueno *et al.* (2011)*	statistical	PCA+kNN, PCA+LBN+SVM	BIRADS	322 MIAS	up to 84%
Tzikopoulos *et al.* (2011)*	fractal dimension+statistical	CARTs, kNN, SVM	Three Categories[1]	322 MIAS	up to 77.02%
He *et al.* (2011)*	Tabár's model based segmentation	kNN	BIRADS, Tabár	320 MIAS	70%; 53%
He *et al.* (2012)*	density based segmentation	kNN	BIRADS, Tabár	321 MIAS	78%; 85%

Three Categories[1] represents three density categories, including fatty, glandular, and dense. Three Categories[2] indicates fibrosis, involution atrophy, and normal. The methods marked with $*$ will be included into a complete comparison with our proposed methods in Chapter 8. The details of the experimental data, evaluation methodologies, and classification results for the various methods will be discussed.

which the area and perimeter of a binary image pattern as well as the number of connected components can be extracted. Discriminant analysis was used to assess the predictive potential of the extracted features.

Subashini *et al.* (2010) extracted fourteen statistical features from the breast region and trained the SVM classifier using different combinations of these features. Finally, nine features including mean, standard deviation, smoothness, skewness, uniformity, kurtosis, average histogram, modified standard deviation, and modified skewness were selected for breast tissue classification with three density classes.

Bueno *et al.* (2011) classified breast parenchymal density according to the BIRADS density categories. Different classification methods were tested and a hierarchical classification procedure was used based on the kNN classifier and a combination of linear Bayes normal (LBN) and support vector machine (SVM). Statistical features were used to characterise the texture of the whole breast, obtained from greylevel histograms and co-occurrence matrices. Moreover, principal component analysis was applied to reduce the feature space dimensionality.

Tzikopoulos *et al.* (2011) used a new fractal dimension related feature and a new set of statistical features for breast density classification. The fractal related feature was computed based on the power spectrum of the segmented breast area. The statistical features were extracted from a set of regions within the breast, segmented by multiple thresholds determined by the minimum cross entropy thresholding (Brink & Pendock, 1996). The mean and the variance of the pixel intensity values were computed for each region. Three different classifiers, classification and regression trees (CARTs), kNN and SVM, were used for the classification of mammograms according to the three density classes (fatty, glandular, and dense).

He *et al.* (2010, 2011, 2012) used the resulting mammographic segmentation to classify mammographic images into BIRADS and Tabár categories. The relative proportions of the segmented regions to the whole breast formed the feature space and the kNN classifier was used for image classification.

In Chapter 6, we will apply the proposed methods to mammographic risk assessment by means of classifying mammographic images into the four BIRADS density categories. This means all the developed methods can extract discriminative and informative features from mammographic images, and the BIRADS breast density classification is the final stage of automated mammographic risk assessment. Hence, we end the literature review by summarising existing work on breast density classification, with particular focus on the publications related to the BIRADS breast density classification. These publications will be used as the baseline references for conducting a comprehensive comparison with our work.

2.5 Summary

In this chapter, we have reviewed a range of existing mammographic image analysis approaches. A brief taxonomy has been presented by dividing these approaches into mammographic image segmentation, mammographic density estimation, mammographic parenchymal patterns characterisation, and breast density classification. For each partition, all the reviewed approaches were summarised in a table, which enables a quick reference to the key techniques in these approaches. As noted at the end of each section above, this literature review has established a basis for our own research performed in this thesis. Our work on mammographic risk assessment will be presented in the subsequent chapters, including breast region segmentation, breast density segmentation, mammographic tissue appearance modelling, and mammographic parenchymal patterns representation.

Specifically, regarding breast region segmentation, although many methods have been developed, it is still a difficult task for digitised mammograms due to the presence of high levels of noise and artifacts. This motivated us to combine a number of well-known image segmentation techniques in order to achieve a comprehensive solution and develop a new breast region segmentation method. Using the developed method, we can identify the breast region for each mammogram, which is an important pre-requisite for the development of a complete framework for automated mammographic risk assessment. As indicated by many studies in the literature, mammographic density and parenchymal patterns are strongly associated with breast cancer risk. Therefore, for the purpose of automated mammographic risk assessment, it is still worth developing effective and robust methods for quantifying mammographic density and/or characterising mammographic parenchymal patterns. As is evident from existing methods, greylevel based features are generally used to describe breast tissue density and texture features are mainly used to represent breast tissue patterns. On the basis of this, we attempt to achieve quantitative estimation of breast density using intensity based approaches. Two breast density segmentation methods are proposed for segmenting the breast into different density sub-regions and detecting dense tissue regions from the breast, respectively. On the other hand, we propose to model mammographic parenchymal patterns using texture representations. A method for modelling mammographic tissue appearance is proposed based on local features. In addition, it would be interesting to develop a new representation of mammographic parenchymal patterns instead of using conventional texture features as in existing approaches. Finally, all the proposed methods should have the potential for automated mammographic risk assessment and breast density classification.

Chapter 3

Image Segmentation

Image segmentation is the process of partitioning an image into regions such that each region is homogeneous with respect to one or more properties or features (such as brightness, colour, texture, reflectivity, etc.) (Pal & Pal, 1993; Gonzalez & Woods, 2002; Sonka *et al.*, 2007). In medical image processing, segmentation plays an important role for automating or assisting the delineation of anatomic structures and the detection of regions of interest (Pham *et al.*, 2000). For mammographic image analysis, image segmentation can be used for detecting masses, microcalcifications and spiculated lesions, and estimating breast density based on the segmented dense tissue regions. A large variety of segmentation techniques have been developed, but there is no single method that can provide good performance for all types of images or applications, nor are all methods equally good for a specific image type or application (Pal & Pal, 1993). In this chapter, we first provide a brief review of the most used image segmentation methods in the literature, introducing the core concepts and discussing the main advantages and disadvantages of the various approaches. On the basis of this, we present our developed segmentation methods for mammographic image analysis. In Section 3.2, a breast region segmentation method is presented. In Sections 3.3 and 3.4, two breast density segmentation methods are presented, which are based on a modified fuzzy c-means algorithm and a topographic approach, respectively.

3.1 A Brief Review of Image Segmentation

A large number of image segmentation approaches exist in the literature. This section provides a brief review of common image segmentation methods covering the following categories: thresholding, edge based segmentation, region based segmentation, clustering, classifier based segmentation, deformable model based segmentation, and some other segmentation methods. A full description of how

these methods are implemented is beyond the scope of this chapter. Here, we briefly introduce the basic concepts of these methodologies and provide some representative references.

3.1.1 Thresholding

Thresholding is the simplest method for image segmentation. When only a single threshold is used for the whole image, it is called global thresholding. If the threshold varies over the image and depends on local characteristics of sub-images, it is called local thresholding. In addition, thresholding can also be divided into bi-level thresholding and multi-thresholding. In bi-level thresholding, the resulting image is a binary image, where pixels with value 1 correspond to objects, and pixels with value 0 correspond to the background. In multi-thresholding, the resulting image is no longer binary, but consists of a set of values corresponding to independent objects. Threshold selection is important for the segmentation result by thresholding. This can be determined interactively based on an operator's visual assessment of the segmentation result. On the other hand, there are many automatic threshold selection methods. Some of them are based on histogram shape analysis, such as detecting peaks and valleys of a histogram (Sonka *et al.*, 2007). Others are based on optimising a criterion function, such as minimising the total misclassification error, maximising a measure of class separability, or maximising a posterior entropy of the partitioned image, where the histogram is assumed as a mixture of two or more normal probability distributions (Pal & Pal, 1993). In addition to using the histogram to determine the threshold, there are some approaches that take into account spatial information based on the co-occurrence matrix of the image. A survey of threshold selection methods is provided in (Weszka, 1978) and an overview of thresholding techniques can be found in (Sahoo *et al.*, 1988).

3.1.2 Edge Based Segmentation

Image segmentation can also be obtained based on detection of edges of various regions in an image. Edge based segmentation is one of the earliest segmentation approaches and is still very important. Edge detection aims to find locations of discontinuities in intensity values, which can be detected using the first-order and second-order derivatives of the image function. In image processing, the first-order derivative can be approximated by the gradient, and the second-order derivative can be computed using the Laplacian. Consequently, on the one hand, many edge detection methods are based on gradient operators, such as Sobel, Prewitt and Roberts (Gonzalez & Woods, 2002). On the other hand, some edge detection

methods use the Laplacian operator to detect zero crossings (Gonzalez & Woods, 2002). All edge detection methods based on the gradient or Laplacian are very sensitive to noise. Some methods reduce noise sensitivity by smoothing the image before applying edge detection. Marr & Hildreth (1980) used a Gaussian filter to smooth the image before applying the Laplacian operator (i.e. Laplacian of Gaussian (LoG)). Canny (1986) proposed an approach to edge detection where the image is also smoothed using a Gaussian filter. The Canny detector is a very powerful edge detector and was deemed to be superior to the Sobel and LoG detectors (Gonzalez *et al.*, 2004).

The resulting image of edge detection cannot be directly used as a segmentation result, as the resulting edge pixels rarely represent an edge completely. Therefore, supplementary processing steps are essential to combine the edge pixels into meaningful edges corresponding to borders of existing objects. To remove false edges resulting from quantisation noise and small illumination irregularities, edge magnitude based thresholding can be applied to the edge image (e.g. non-maximal suppression). Moreover, some approaches aim to construct continuous borders: the Hough transform (Hough, 1962) finds and links line segments; edge relaxation repairs missing parts of borders based on edge context in the local neighbourhood; and border tracing tracks a border starting at a pixel and finds subsequent border elements based on edge magnitudes and directions (Sonka *et al.*, 2007).

3.1.3 Region Based Segmentation

In the previous section, image segmentation was obtained by finding borders between regions. In contrast, the segmentation methods to be described in this section find the regions directly. For region based segmentation methods, homogeneity is an important property and is used as the main segmentation criterion, which can be based on greylevel, colour or texture. Here, we introduce three basic region based segmentation techniques: region growing, split and merge, and watershed methods.

Region growing, the most well-known region based method, is a procedure of grouping pixels into regions based on a pre-defined growing criterion. The growing procedure starts with a seed point or a set of seed points, and a region is iteratively grown by adding neighbouring pixels that have similar properties to the seed point. This procedure stops when no more neighbouring pixels satisfy the growing criterion based on the homogeneity of the grown region (Gonzalez & Woods, 2002). The seed point can be manually selected by an operator or placed using an automatic seed finding step, and the homogeneity criterion depends on the particular application and the image type (such as using colour information

for satellite imagery and using intensity information for X-ray radiography).
As the name implies, the split and merge method firstly subdivides an image into a set of sub-regions and then merge these sub-regions to obtain the final segmentation. The initial splitting process recursively splits the image until all regions satisfy a homogeneity criterion, and the accompanying merging step merges adjacent regions that satisfy the pre-defined constraints (Gonzalez & Woods, 2002).
The watershed transform is a very powerful segmentation tool based on morphology (Soille, 2003), which applies the concepts of *watershed* and *catchment basin* in geography to realise image segmentation. A grey-scale image is regarded as a topological surface, where the intensity value at each pixel is interpreted as the height. If water falls on this surface, it will be collected into the *catchment basins*, and the water level of each *catchment basin* will rise until reaching a ridge (*watershed*) which separates two adjacent *basins*. Thus, the watershed transform segments the image by finding the *catchment basins* and ridges on the corresponding surface.
Region based segmentation techniques generally perform better than edge based methods in noisy images, where borders between regions are extremely difficult to detect (Sonka *et al.*, 2007). However, region based segmentation can also be sensitive to noise, causing the resulting regions to have holes or be disconnected (under-segmentation). Conversely, over-segmentation can be caused by partial volume effects where separate regions are connected (Pham *et al.*, 2000). A combination of edge based methods and region based methods tends to provide better segmentation results (Sonka *et al.*, 2007).

3.1.4 Clustering

Clustering is the process of classifying observations (data items, patterns, or feature vectors) into groups (called clusters) (Jain *et al.*, 1999), which can be unsupervised learning aiming to find structure in an unlabelled dataset. Observations in the same cluster are similar in some way and are dissimilar to those belonging to other clusters. Clustering algorithms can also be used for image segmentation by assigning every pixel to one cluster in the feature space, which iterate between partitioning pixels and computing properties of each cluster. There are three commonly used clustering algorithms: k-means, fuzzy c-means, and expectation maximisation algorithms.
The k-means (Mac Queen, 1967) is the simplest clustering algorithm for partitioning a dataset into k subsets. It starts with an initial partition and iteratively updates the partition based on the similarity between data points and cluster centres until a convergence criterion is satisfied, such as no reassignment of any data point, or the objective function ceases to decrease significantly after some

number of iterations. Although this iterative procedure will always terminate, the k-means algorithm generally will not necessarily converge to an optimal partition (i.e. the global minimum of the objective function). Moreover, the algorithm is significantly sensitive to the initial partition and is more likely to converge to a local minimum resulting from improper initialisation. However, the k-means is the most popular clustering algorithm and has been widely used due to its easy implementation and reasonable time complexity (Jain *et al.*, 1999).

The fuzzy c-means algorithm incorporates fuzzy set theory (Zadeh, 1965) into the k-means clustering, which allows one data point to belong to two or more clusters. It was initially developed by Dunn (1973) and improved by Bezdek (1981). The details of the standard fuzzy c-means algorithm and its variations will be provided in Section 3.3.1.

The expectation maximisation (EM) algorithm is also an iterative method, which aims at finding maximum likelihood estimates of parameters in statistical models. It was proposed by Dempster *et al.* (1977) and the detailed description can be found in (Bishop, 2006). The EM procedure starts with an initial estimation of the parameters in a mixture of models and iterates between an expectation step and a maximisation step: the expectation step computes the expected value of the likelihood function, and the maximisation step updates the parameter estimation to maximise the expected likelihood value. It has been effectively applied to Gaussian mixture modelling (Redner & Walker, 1984; Rasmussen, 2000), where a number of Gaussian components are used to fit a multimodal distribution. The means and covariances of the Gaussian components can be estimated and updated in the EM iteration, and the resulting multiple Gaussians are combined to form a generalised model. This algorithm is also very sensitive to initialisation.

Although clustering algorithms can be unsupervised learning, an initialisation step is always essential in the clustering procedure. An inappropriate initialisation can significantly affect the resulting segmentation. In addition, the clustering algorithms do not always directly incorporate spatial information of neighbouring pixels and therefore are sensitive to noise and intensity inhomogeneity.

3.1.5 Classifier Based Segmentation

Image segmentation can also be obtained using classifiers that assign a label to every pixel. Classifier based segmentation methods use pattern recognition techniques that attempt to partition feature vectors extracted from an image into a given set of classes, which tend to be supervised methods where features derived from reference segmentations (e.g. manual segmentation) are used as training data. The features used in the classification procedure are varied, which can be

related to intensity, texture, or other properties of the image.
A simple classifier is the k-Nearest Neighbours classifier, where each object is classified based on a majority vote of its k closest training samples. In the sense of image segmentation, each pixel is labelled with the most common class among the k nearest neighbours of its corresponding feature vector in the feature space. A detailed introduction of the algorithm can be found in (Duda & Hart, 1973). It has been applied to medical image segmentation, such as pulmonary nodule detection in chest CT (Murphy *et al.*, 2009) and brain tissue segmentation in MR data (Vrooman *et al.*, 2007).
Another simple classifier is the naive Bayes classifier, which is a simple probabilistic classifier based on the naive Bayes assumption where features are independent within a given class (Duda & Hart, 1973; Bishop, 2006). Although independence is generally a poor assumption, the naive Bayes classifier can often perform as well as more sophisticated classifiers (Rish, 2001). It has proven to be effective in many real applications, such as text segmentation/classification and texture image segmentation (Hsiao & Sawchuk, 1989; Haji & Katebi, 2005).
In addition, a number of segmentation methods are based on a neural network classifier (Pal & Pal, 1993). Neural networks are composed of a large number of highly connected processing elements (neurones) working simultaneously to deal with specific problems. An artificial neural network is constructed following the way of biological information processing systems (such as the brain), and is trained for specific applications through a learning process (Bishop, 1995; Ripley, 1996). In the last two decades, there has been a wide application of neural networks in image processing, such as image segmentation and object recognition (Egmont-Petersen *et al.*, 2002).
Classifier based methods exclude the iterative procedure in clustering methods (except for the training phase of the classifier), and therefore they are relatively computationally efficient. However, the interactive/manual segmentation for obtaining training data can be time consuming and labourious. Moreover, for the methods in which high-dimensional features are used, the pixel-wise classification can also involve high computational time. On the other hand, using a single training set for a large variety of images can cause biased results, this indicates a lack of flexibility of the classifier based methods. As such, the classifiers will have to be retrained if the training set is updated.

3.1.6 Deformable Model Based Segmentation

Deformable models are model based methods for delineating boundaries of objects using deformable curves or surfaces, which have been widely used in image

segmentation, especially for segmenting anatomic structures in medical images. Deformable curves or surfaces are defined within a domain of a 2D or 3D image that can deform under the influence of internal and external forces. The internal forces are derived from the curve or surface itself to keep it smooth during the deformation. The external forces are computed based on the image data to move the curve or surface towards the object boundary (Xu *et al.*, 2000). By incorporating a prior knowledge of the object shape and constraining the resulting boundary to be continuous and smooth, deformable models can compensate for image noise and boundary gaps or irregularities and can provide a complete and consistent mathematical description for the object boundary (McInerney & Terzopoulos, 1996; Xu *et al.*, 2000).

The idea of deforming a model for image analysis originated a long time ago and has become popularly known since "snakes" models were proposed (Kass *et al.*, 1988). Deformable models can generally be categorised into two main groups: parametric deformable models (Cohen, 1991; McInerney & Terzopoulos, 1995) and geometric deformable models (Caselles *et al.*, 1993; Malladi *et al.*, 1995). Parametric deformable models represent curves and surfaces explicitly in a parametric form, including two types of formulations: an energy minimising formulation and a dynamic force formulation. Although they can provide promising results for many segmentation tasks, they may yield cusps or intersecting boundaries in some situations (Sonka *et al.*, 2007). In contrast, geometric deformable models represent curves and surfaces implicitly as level sets of higher dimensional functions. Curves and surfaces are evolved only using geometric computations, allowing topological changes to be easily accommodated (Xu *et al.*, 2000; Sonka *et al.*, 2007).

In medical image analysis, deformable models have proven to be effective in segmenting, matching, and tracking anatomic structures, with the capability of accommodating the large variability of anatomic structures over time and across different individuals (McInerney & Terzopoulos, 1996). However, a possible disadvantage of deformable models is that they may need manual interaction to initialise the model and choose the appropriate parameters. Moreover, standard deformable models can show an inappropriate convergence to certain types of boundaries (e.g. concave contours) (Pham *et al.*, 2000).

3.1.7 Other Methods

Some methods use Markov random field theory (Li, 2009) to model spatial interactions between neighbouring pixels for image segmentation in the sense that most pixels belong to the same class as their neighbouring pixels. Zhang *et al.* (2001) combined the hidden Markov random field model and the EM algorithm

to obtain brain MR image segmentation.

Atlas guided approaches can also be used for image segmentation. A method based on an anatomic pulmonary atlas was developed to segment the oblique lobar fissures in volumetric X-ray CT images (Zhang *et al.*, 2006). In atlas guided approaches, an atlas is first generated by assembling information on objects in images to be segmented. This atlas is then used as a reference to segment unseen images. Atlas guided approaches are to some extent similar to classifier based methods, while they are straightforward implemented in the spatial domain of the image rather than in the feature space (Pham *et al.*, 2000).

Some segmentation methods use fuzzy connectivity to describe the segmentation task with fuzzy rules instead of crisp relationships (Sonka *et al.*, 2007). Fuzzy connectivity provides the flexibility for distinguishing between object and non-object regions, which can often be a powerful solution to difficult segmentation problems (Udupa & Samarasekera, 1996).

Mean shift has also been applied to image segmentation, which is a non-parametric iterative algorithm for estimating the density gradient and seeking the mode of a given distribution (Cheng, 1995). Image segmentation can be obtained by applying the algorithm in the joint spatial-range domain (Comaniciu & Meer, 1999).

In addition, there are some methods based on graph theory, by transforming the image segmentation problem into a graph based representation, where each node corresponds to a pixel in the image and an edge connects two nodes with a weight based on some property (e.g. intensity) of the two corresponding pixels (Felzenszwalb & Huttenlocher, 2004).

In practical applications, combined strategies are often used to deal with a specific segmentation problem, by simultaneously combining a set of segmentation methods described above rather than using them individually (Haris *et al.*, 1998). On the other hand, some segmentation methods are based on integration of information from multiple images, such as multi-spectral or multi-modal images (Reddick *et al.*, 1997; Farag *et al.*, 2006). There are many other methods for image segmentation besides the approaches described above, but a detailed literature review is out of the scope of this thesis.

3.2 Breast Region Segmentation in Mammograms

As stated in Section 2.1.1, breast region segmentation is an essential pre-processing step in the (semi-)automatic analysis of mammographic images. It separates the background and the pectoral muscle from breast tissue to avoid interference in further processing. The precise segmentation of the breast region with a minimum of breast tissue pixels mis-segmented is important for searching for abnormalities,

modelling parenchymal patterns, and accurate registration. We develop a breast region segmentation method by taking and combining a number of segmentation approaches described in the previous section, including histogram thresholding, edge detection, active contour modelling, polynomial fitting, and region growing.

3.2.1 Methodology

An example mammographic image (5671 × 3788 pixels) taken from the EPIC database (Day *et al.*, 1999) is shown in Figure 3.1(a). This image contains left and right MLO mammograms and needs to be split into individual mammograms. Firstly, a global thresholding is used for approximate segmentation. The threshold, after Gaussian smoothing of the histogram, is determined using the valley between the two peaks which correspond to the background and breast tissue, respectively (Figure 3.1(b)). The resulting binary image contains a number of objects due to the non-uniform background (Figure 3.1(c)). We use a connected component labelling (8-connected neighbourhood) algorithm (Davies, 1997) to remove labels and annotations in the background region (Figure 3.1(d)). Subsequently, we isolate the frame (near the edges of the image) and smooth the remaining region using a Gaussian filter. Finally, we split the union region containing bilateral breasts into two separate breast regions to form two binary masks (Figure 3.1(e)). The split left and right MLO mammograms superimposed with the corresponding mask boundaries are shown in Figure 3.1(f).

The approximate segmentation is refined using scale-space based edge detection. The right MLO mammogram in Figure 3.1(f) is used as an illustration. Firstly, we evenly place 40 points on the mask boundary (Figure 3.2(a)). For each point an orthogonal line is obtained (Figure 3.2(b)). The length of orthogonal lines is 500 pixels (100 pixels inside the mask and 400 pixels outside the mask). We then perform edge detection to search for probable breast boundary points by convolving the pixels on orthogonal lines with a derivative of Gaussian kernel at multiple scales (Canny, 1986). We use a range of small scales in order to increase sensitivity to the low contrast breast boundary. Edge detection starts at the coarsest scale within the scale range to suppress noise, and ends at the finest scale to improve accuracy. The probable breast boundary points are achieved by detecting the minima of the convolution results. Figure 3.3 indicates the procedure of finding the probable boundary points on one orthogonal line (parallel to the x-axis direction) across multiple scales (from coarse to fine scales). Figure 3.3(a) shows the intensity profile of the orthogonal line (which is the starting line of contour growing in the following paragraph), which indicates the lack of a distinct edge for the breast boundary. Figures 3.3(b), 3.3(c), and 3.3(d) show the convolution

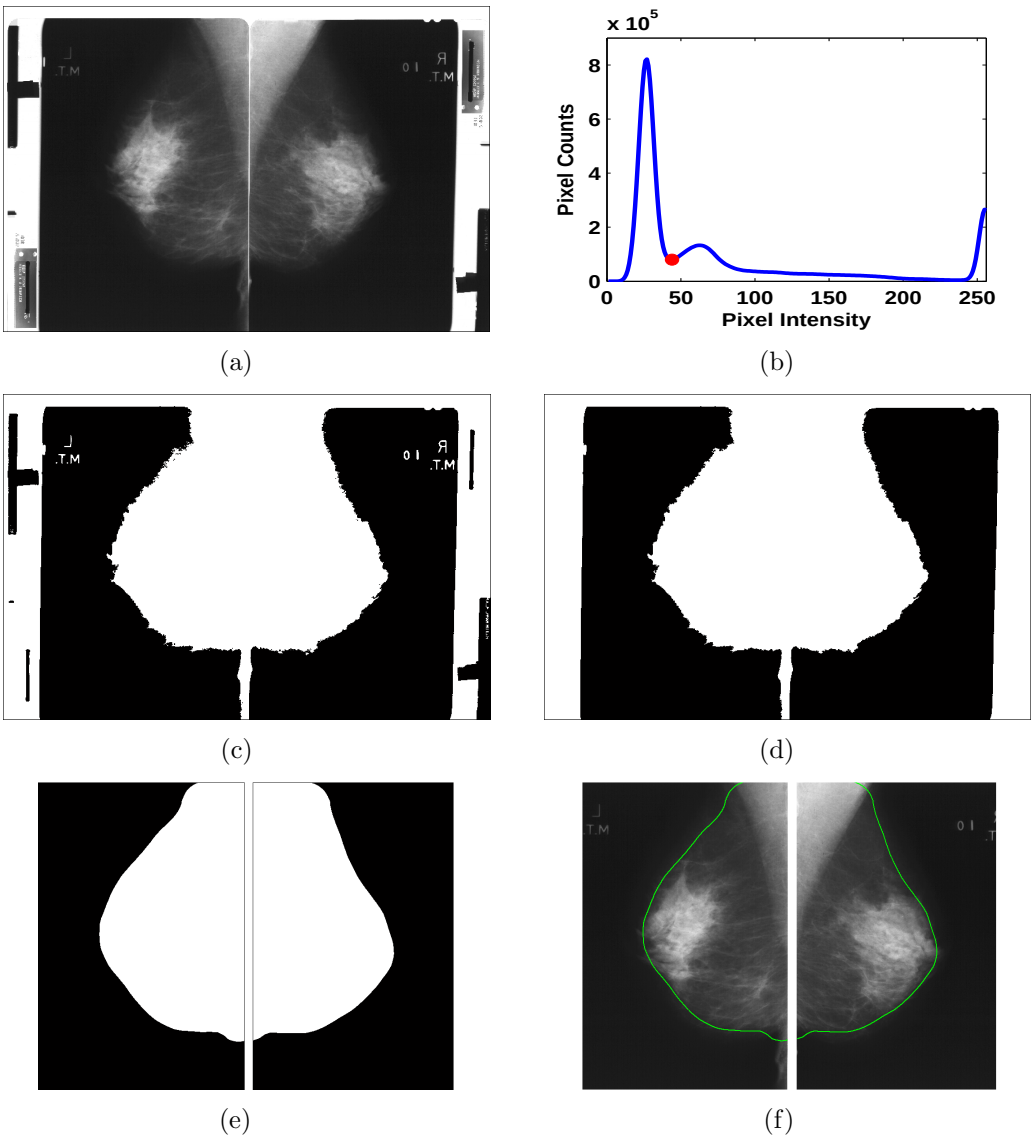


Figure 3.1: Approximate segmentation: (a) mammographic image; (b) intensity histogram and selected threshold; (c) resulting binary image; (d) labels and annotations removal; (e) binary masks of bilateral breasts; and (f) mask boundaries.

results at three scales, where the detected edge points (i.e. probable boundary points) at each scale are marked with red stars. The resulting probable breast boundary points on each orthogonal line are shown in Figure 3.2(c).

An active contour model is used to generate a preliminary breast boundary from the resulting probable breast boundary points. The first step of contour growing is identifying a starting orthogonal line and selecting a seed point from all the probable breast boundary points on this line. We give priority to choosing the orthogonal line close to the x-axis direction as the starting line. We use an edge strength measure to search for the seed point along the starting line from the mask towards the background. Ideally, the seed point can be located at the boundary point where the first local maximal edge strength is obtained (the circled boundary

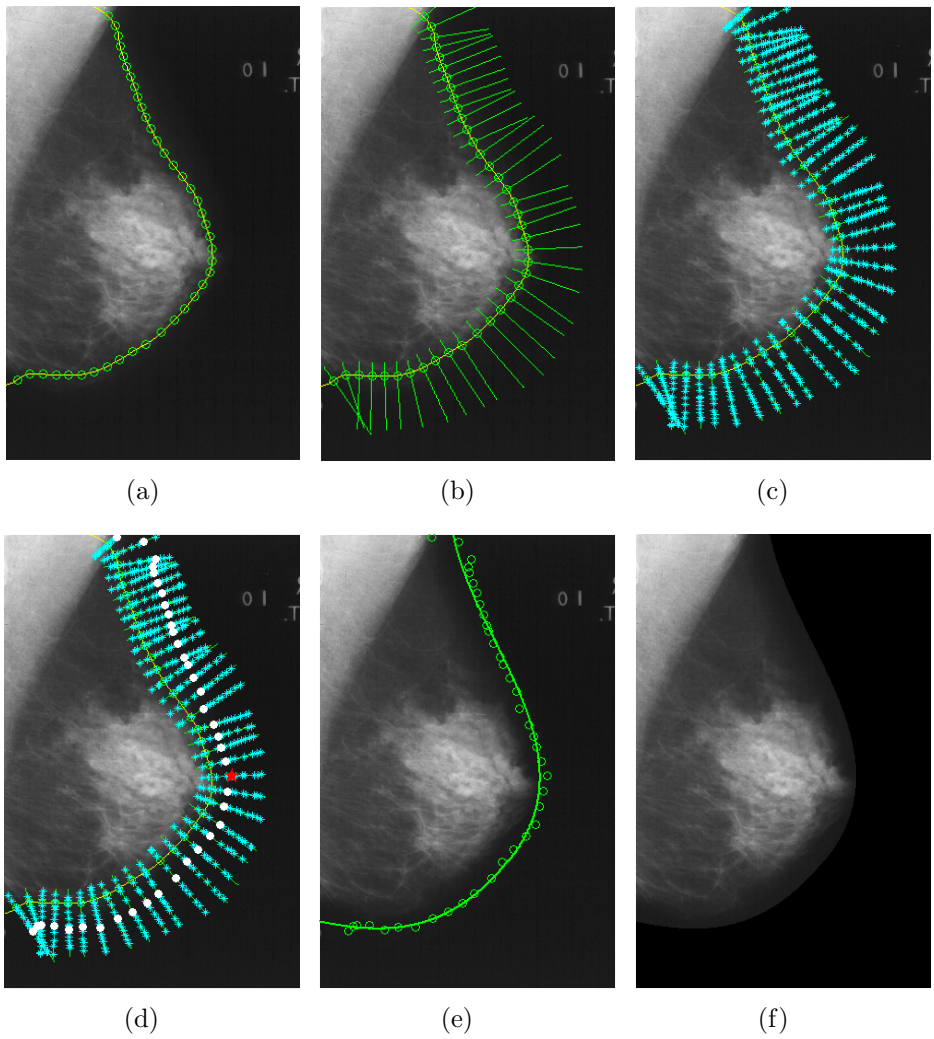
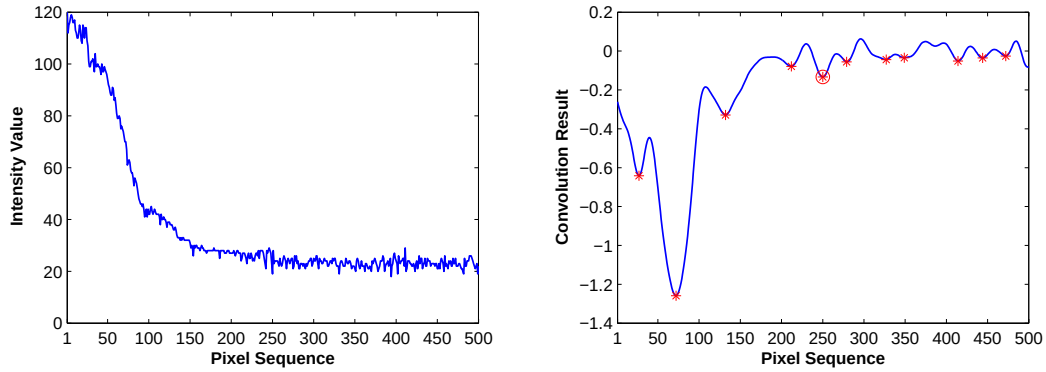


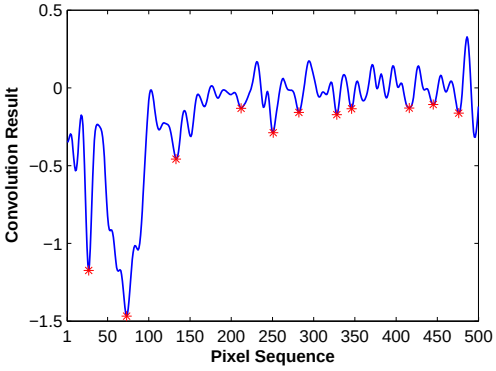
Figure 3.2: Overview of edge detection in scale space: (a) place 40 points on mask boundary; (b) plot an orthogonal line for each point; (c) detect probable breast boundary points; (d) initial breast boundary (white circles) and the first seed point (red star); (e) final breast boundary; and (f) background removal.

point as shown in Figure 3.3(b)). If no such a seed point exists on this starting line, other alternatives close to it will be used to start a new searching procedure dynamically. After the first seed point is selected, a contour growing process starts based on a contour growing measure, which is defined by a weighted function of intensity, edge strength and angle information, following the typical snake additive model formulation (Martí *et al.*, 2007). For a seed point, the probable breast boundary points on the neighbouring orthogonal line are regarded as candidates for a new seed point. The contour growing measure is calculated for all candidate points to decide the new seed point which has the minimum measure value. Once 40 seed points are obtained contour growing is finished, and these 40 seed points comprise an initial breast boundary (Figure 3.2(d)). We then order these points to avoid the possible disorder due to the intersection of the orthogonal lines and

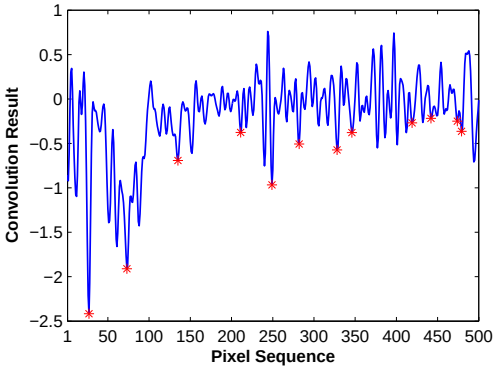


(a) intensity profile

(b) $\sigma = 8$



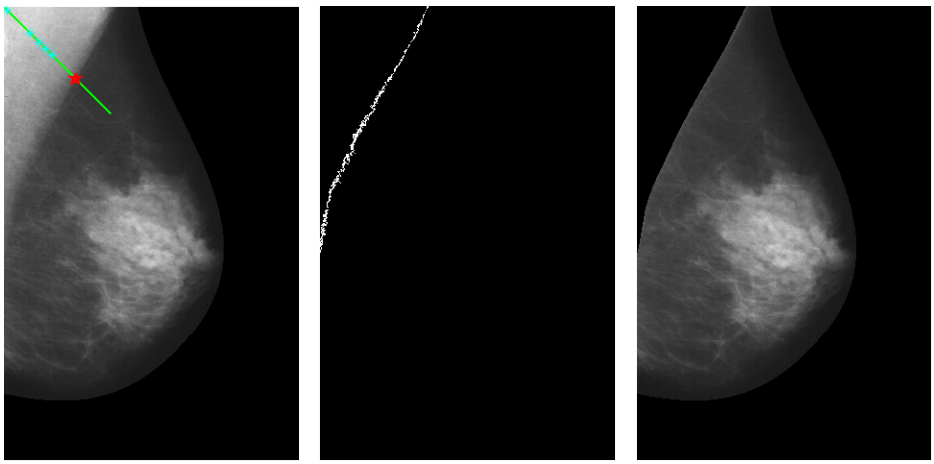
(c) $\sigma = 4$



(d) $\sigma = 2$

Figure 3.3: Multiscale edge detection: (a) intensity profile of one orthogonal line; (b) convolution result when $\sigma = 8$ and the position of minima is 27, 72, 132, 212, 250, 279, 327, 349, 414, 444 and 472; (c) convolution result when $\sigma = 4$ and the position of minima is corrected to 27, 73, 133, 212, 251, 282, 328, 346, 416, 445 and 476; and (d) convolution result when $\sigma = 2$ and the position of minima is corrected to 27, 73, 135, 211, 249, 282, 328, 346, 419, 442 and 474.

combine close points into one point. Finally, a cubic polynomial fitting is used to yield a smooth and continuous contour as the breast boundary (Figure 3.2(e)). The resulting breast region after removing the background is shown in Figure 3.2(f). The breast region obtained above is the union region of the breast and the pectoral muscle (Figure 3.4(a)). We use a region growing method to remove the pectoral muscle. Firstly, we place a seed point close to the border between the pectoral muscle and breast tissue instead of placing a seed point inside the pectoral muscle region (Raba *et al.*, 2005). Specifically, we draw a line (slope equal to 1 or -1) from the first pixel of the non-curved side into the breast, and then we detect edges on this line using the same method as described earlier. The seed point is selected from these detected edge points using a measure incorporating aspects of edge strength and edge position (Figure 3.4(a)). After that, a region is grown from the seed point based on the similarity with the region's mean intensity. In traditional region growing, the region is iteratively grown until the intensity



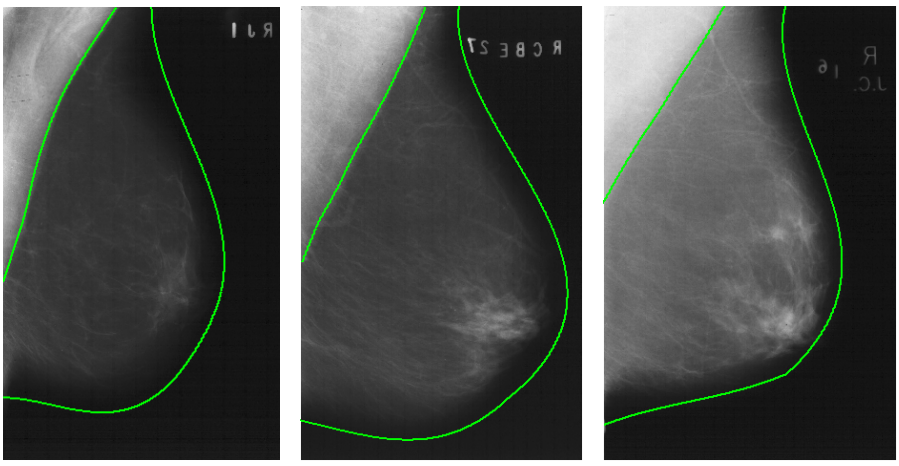
(a) (b) (c)

Figure 3.4: Pectoral muscle removal: (a) seed point of region growing (red star); (b) resulting region of region growing; and (c) resulting breast region.

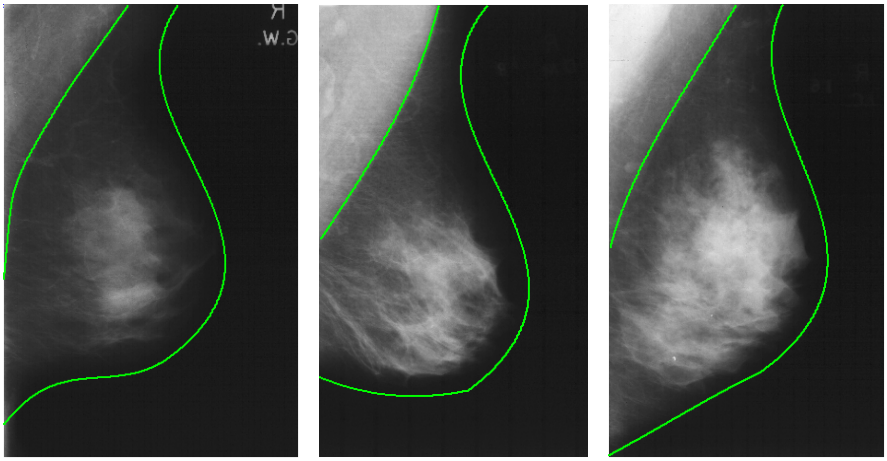
difference between the region's mean and new neighbouring pixel is larger than a threshold. Here, we use a new termination criterion to efficiently avoid under-segmentation of inhomogeneous regions. Region growing starts with a critical initial threshold of intensity difference, and the threshold iteratively increases in the growing process. This growing process stops when the resulting region nearly approaches the edges of the image (Figure 3.4(b)). Finally, we use locally weighted scatterplot smoothing (Cleveland & Devlin, 1988) to refine the pectoral muscle boundary, which can accurately preserve the boundary feature. The resulting breast region with the pectoral muscle removal is shown in Figure 3.4(c).

3.2.2 Results and Discussion

This method has been tested using 248 MLO mammograms of the EPIC database and 321 MLO mammograms of the MIAS database (mdb295ll was excluded for historical reasons). To demonstrate the validity of this method for segmenting mammograms with different breast density categories, segmentation results of example mammograms with six density categories (SCC1 to SCC6) are shown in Figure 3.5. All segmentation results were evaluated by two scientists in our research group (both have an image processing background). They were visually rated into four categories: *accurate*, *nearly accurate*, *acceptable* and *unacceptable*. The four categories were the commonly used rating forms for the evaluation of breast region segmentation algorithms in the previously published papers (Bick *et al.*, 1995; Méndez *et al.*, 1996; Ojala *et al.*, 2001). *Accurate* or *nearly accurate* was rated according to whether the segmentation result preserved the breast/pectoral muscle boundary exactly or nearly exactly. Otherwise, the result was rated as



(a) SCC1 (b) SCC2 (c) SCC3



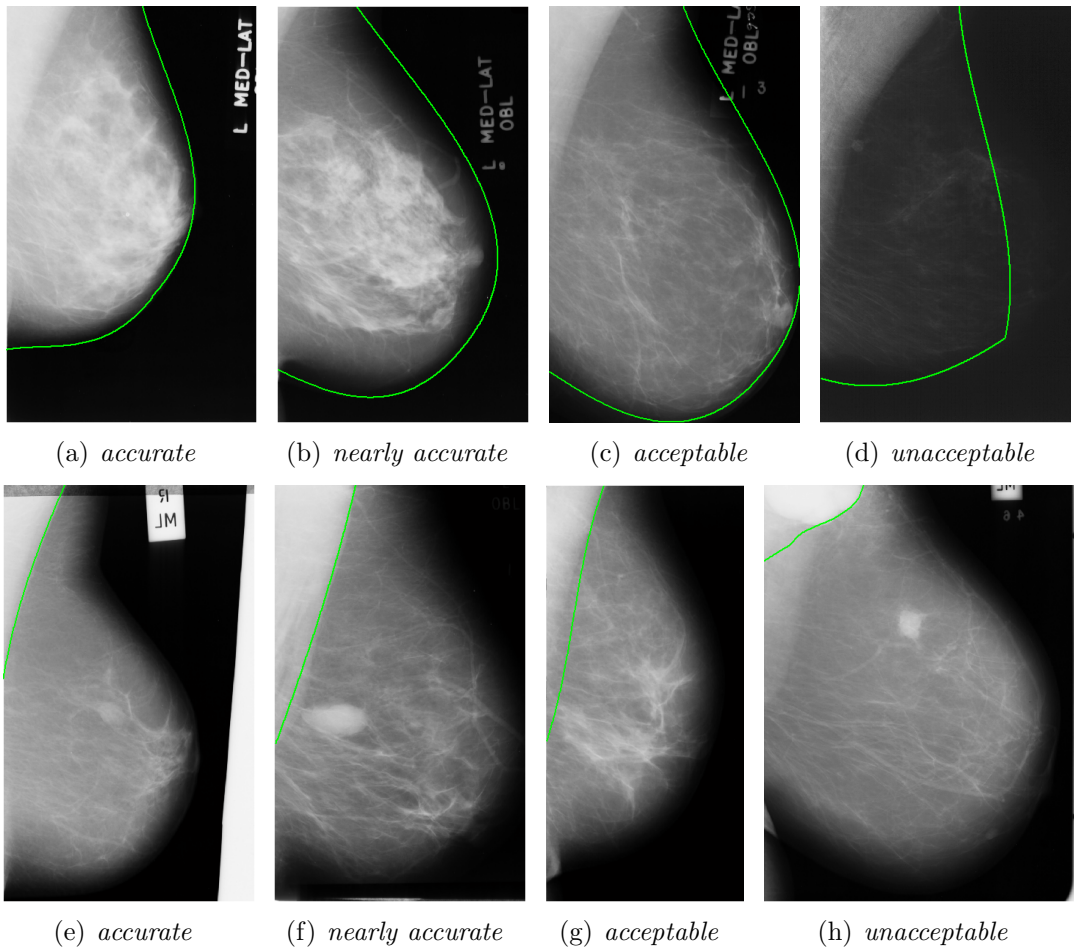
(d) SCC4 (e) SCC5 (f) SCC6

Figure 3.5: Segmentation results of example mammograms with six density categories ranging from SCC1 to SCC6.

Table 3.1: Segmentation results for the EPIC and MIAS databases. A, B, C and D denote the four rating categories: *accurate*, *nearly accurate*, *acceptable* and *unacceptable*, respectively.

	Breast-Background				Pectoral Muscle			
Database	A	B	C	D	A	B	C	D
EPIC	66.5%	25.0%	6.9%	1.6%	62.5%	25.4%	5.6%	6.5%
MIAS	64.8%	34.0%	1.2%	0.0%	67.9%	24.9%	5.0%	2.2%

acceptable if minor pixels near the boundary were mis-segmented, as those pixels are not relevant and do not provide significant information for CAD purposes. The result with a large deviation from the boundary was rated as *unacceptable*. The obtained segmentation results for the EPIC and MIAS databases are summarised in Table 3.1. Examples of the four categories are provided in Figure 3.6 ranging from *accurate* to *unacceptable* (more examples of *unacceptable* results will be provided in the subsequent discussion).



(a) *accurate* (b) *nearly accurate* (c) *acceptable* (d) *unacceptable*

(e) *accurate* (f) *nearly accurate* (g) *acceptable* (h) *unacceptable*

Figure 3.6: Examples of the four rating categories for the breast-background segmentation (top row) and the pectoral muscle segmentation (bottom row).

We compared our segmentation results with related publications mentioned in Section 2.1.1 where the results were also visually evaluated according to a similar rating criterion for use in CAD schemes. A summary of the comparison is provided in Table 3.2. For the breast-background segmentation, Bick *et al.* (1995) tested their algorithm on 740 digitised mammograms. In 97% of the mammograms, the segmentation results were rated as *acceptable*, and in 82% of the mammograms as *nearly accurate*. Méndez *et al.* (1996) tested their algorithm on 156 digitised mammograms, and the breast segmentation results were rated to be *accurate* or *nearly accurate* in 89% of the mammograms. Chandrasekhar & Attikiouzel (2000) obtained about 94% *acceptable* segmentation results for approximately 300 images from the MIAS database. Ojala *et al.* (2001) evaluated their methods on 20 mammograms digitised with two different digitisers. Ten were randomly selected from the DDSM database, and the other ten were digitised using a Pinja M6000B CCD scanner. The percentages of *acceptable* and *accurate* results were 90% and 55%, respectively. Raba *et al.* (2005) tested their method on all the images from the

Table 3.2: Comparison of our segmentation results with related publications.

	Method	Test Data	*accurate*	*nearly accurate*	*acceptable*	*unacceptable*
Breast-Background	Bick *et al.* (1995)	740		82%	97%	
	Méndez *et al.* (1996)	156		89%		
	Chandrasekhar & Attikiouzel (2000)	300 MIAS			94%	
	Ojala *et al.* (2001)	20	55%		90%	
	Raba *et al.* (2005)	322 MIAS		98%		
	Ours	248 EPIC	66.5%	91.5%	98.4%	1.6%
	Ours	321 MIAS	64.8%	98.8%	100.0%	0.0%
Pectoral Muscle	Karssemeijer (1998)	125			97%	
	Kwok *et al.* (2004)	322 MIAS		83.9%		
	Raba *et al.* (2005)	322 MIAS		86%		
	Ours	248 EPIC	62.5%	87.9%	93.5%	6.5%
	Ours	321 MIAS	67.9%	92.8%	97.8%	2.2%

MIAS database and they obtained 98% *nearly accurate* results. Compared with these methods, we obtained 98.4% *acceptable* and 91.5% *nearly accurate* results for 248 mammograms from the EPIC database, and 100% *acceptable* and 98.8% *nearly accurate* results for 321 mammograms from the MIAS database. For the pectoral muscle segmentation, Karssemeijer (1998) applied the developed method to 615 mammograms from the breast cancer screening programme in Nijmegen. 97% *acceptable* results were obtained in a subset of 125 mammograms. Kwok *et al.* (2004) tested their algorithm on the full MIAS database. 83.9% of the segmentation results were rated to be *adequate* or better. Raba *et al.* (2005) obtained 86% good segmentations for the full MIAS database. By contrast with these methods, we obtained 93.5% *acceptable* and 87.9% *nearly accurate* results for the EPIC database and 97.8% *acceptable* and 92.8% nearly accurate results for the MIAS database. As can be seen from Table 3.2 our results are comparable to or better than the various approaches.

Note that the above is a qualitative comparison, as various approaches were tested using different images and the results were evaluated by different expert mammographic radiologists. In order to further evaluate the validity of our method, we also made a direct comparison of the segmentation results with two methods based on the MIAS database. The first method (Blot & Zwiggelaar, 2001) uses a global thresholding and the Hough transform to remove the background and separate the pectoral muscle from breast tissue as described in (Karssemeijer, 1998). The second method (Oliver *et al.*, 2010) uses the approach developed by Martí *et al.* (2007) to identify the breast region and uses the approach of Kwok *et al.* (2004) to remove the pectoral muscle. We refer to these two methods as Method I and Method II. Segmentation results of four example mammograms from the MIAS database for the three different methods are shown in Figure 3.7. The four mammograms are in different density categories from BIRADS I to BIRADS IV. As

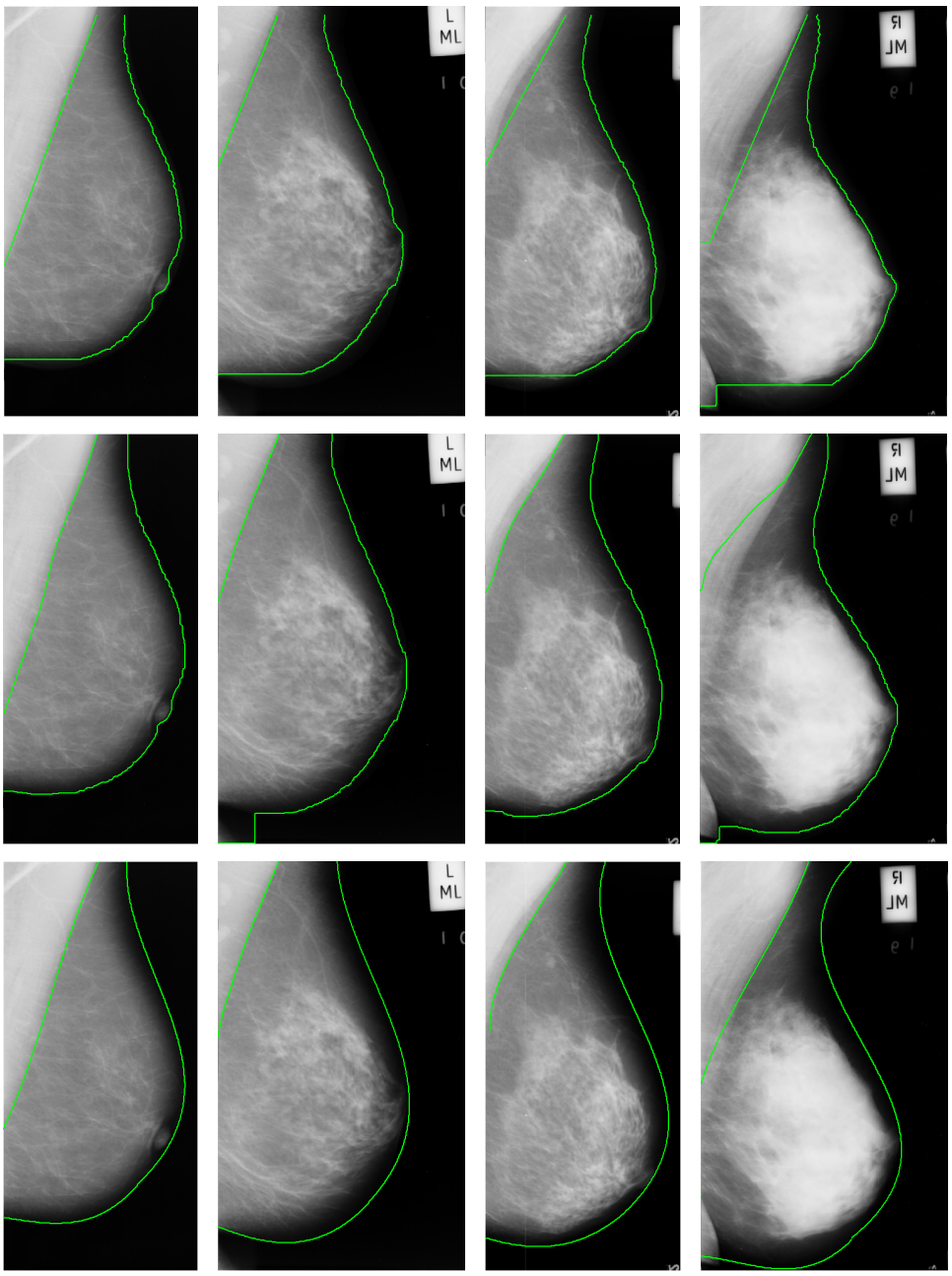


Figure 3.7: Segmentation results of four example mammograms from the MIAS database for the three different methods (top row: Method I, middle row: Method II, and bottom row: our method). The density category of the four mammograms ranges from BIRADS I to BIRADS IV.

shown in Figure 3.7, Method I and Method II tend to under-segment the breast region such that some pixels are lost near the breast boundary, while our method appears to preserve the boundary better than the other methods. Method I uses a straight line to segment the pectoral muscle, which makes it difficult to match a curved boundary. Table 3.3 shows a summary of the comparison, which indicates that our method provides better segmentation results for both the breast-background segmentation and the pectoral muscle segmentation than the other

Table 3.3: Comparison of the segmentation results for 321 mammograms from the MIAS database. A, B, C and D denotes the four rating categories: *accurate*, *nearly accurate*, *acceptable* and *unacceptable*, respectively.

	Breast-Background				Pectoral Muscle			
Method	A	B	C	D	A	B	C	D
Method I	8.7%	57.0%	15.3%	19.0%	45.5%	41.7%	10.0%	2.8%
Method II	51.7%	43.9%	4.1%	0.3%	60.8%	18.7%	12.4%	8.1%
Our	64.8%	34.0%	1.2%	0%	67.9%	24.9%	5.0%	2.2%

two methods. Method I performs worst among the three methods especially for the breast-background segmentation. This is due to the limitations of using a global thresholding for identifying the breast boundary and assuming the pectoral muscle boundary to be a straight line. 97.2% *acceptable* results are obtained for the pectoral muscle segmentation, which is very close to the *acceptable* percentage of 97% obtained by Karssemeijer (1998). Method II generally performs better than Method I. For the pectoral muscle segmentation, the percentage of *nearly accurate* results is 79.5% that is slightly lower than the *adequate* percentage of 83.9% obtained by Kwok *et al.* (2004). However, it should be noted that *adequate* is the third rating in a five-score assessment and it is not necessarily equivalent to the second rating of *nearly accurate* in a four-score evaluation.

In some cases (1.6% of the breast-background segmentations and 6.5% of the pectoral muscle segmentations for the EPIC database, and 2.2% of the pectoral muscle segmentations for the MIAS database) the method failed to obtain what could be considered to be an *acceptable* segmentation. For the *unacceptable* breast-background segmentations, those were mainly caused by an inaccurate binary mask resulting from the extremely low contrast between breast tissue and the background (Figures 3.8(a), 3.8(b) and 3.8(c)). Furthermore, the severe noise in the digitised mammographic images and the serious lack of a distinct edge near the breast boundary could lead to a poor placement of the first seed point for contour growing and yield over-segmented results (Figures 3.8(d) and 3.8(e)). For the pectoral muscle removal, the fuzzy edge between the pectoral muscle and breast tissue could lead to over-segmented results (Figures 3.8(f) and 3.8(g)). Moreover, the layered or inhomogeneous pectoral muscle formed in the mammogram acquisition process could have strong edges within the pectoral muscle region, which penalise the accurate selection of a seed point for region growing and produce under-segmented results (Figures 3.8(h), 3.8(i), 3.8(j) and 3.8(k)).

Future work could focus on further evaluating our method using full-field digital mammograms and improving *unacceptable* segmentation results. In this method, the binary mask plays an important role in subsequent steps, however, it is formed

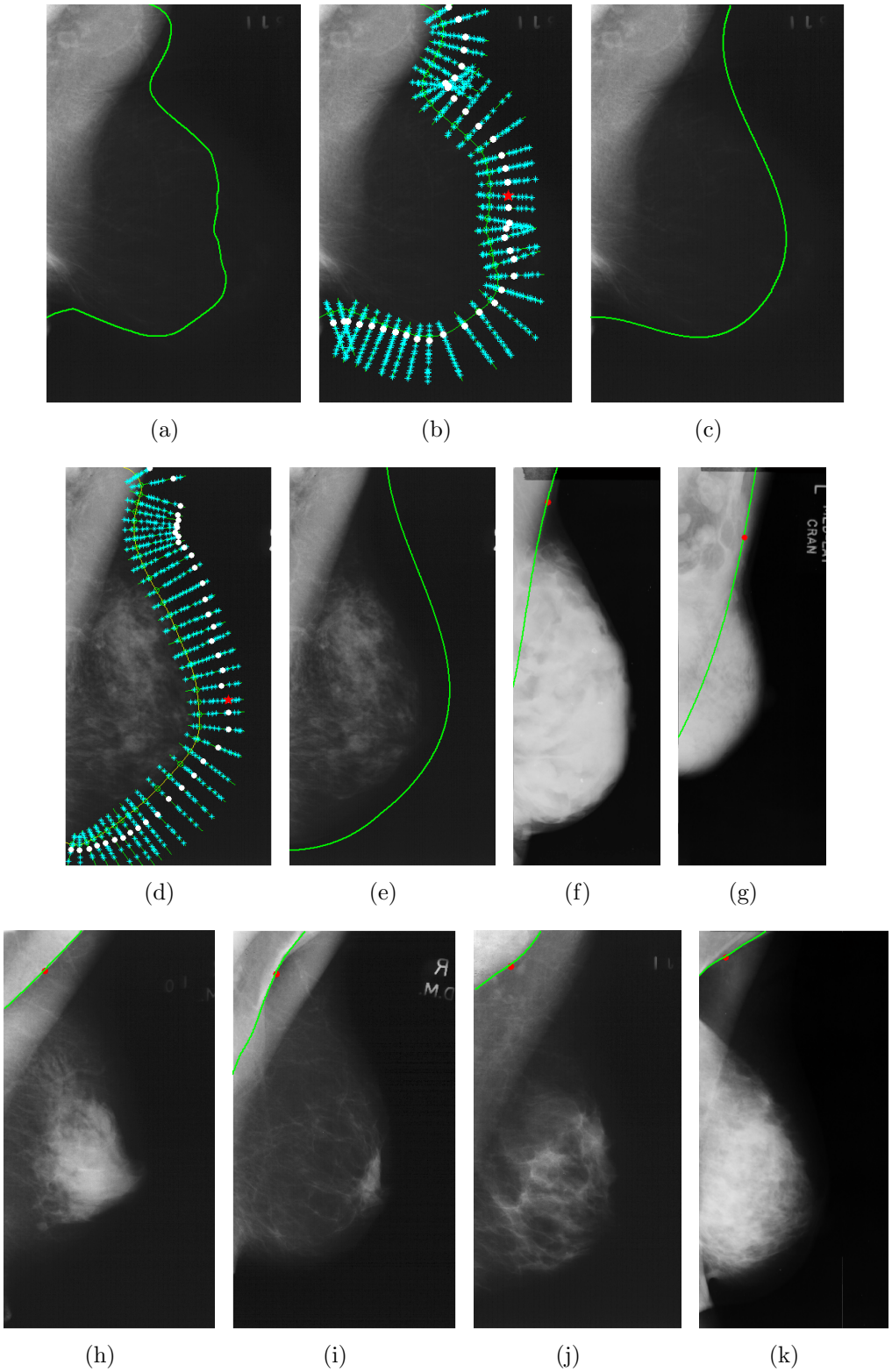


Figure 3.8: Examples of *unacceptable* segmentations: (a), (b) and (c) under-segmentation due to an incorrect binary mask; (d) and (e) over-segmentation due to a misplaced seed point; (f) and (g) over-segmentation due to the mixture of the pectoral muscle and breast tissue; (h), (i), (j) and (k) under-segmentation due to the inhomogeneity of the pectoral muscle region.

using a simple global thresholding directly. Other advanced approaches could be investigated to generate a more accurate binary mask as the approximate segmentation. The weighting factors of intensity, edge strength and regularity terms in the contour growing measure are determined empirically. The influence of these factors on the segmentation performance could be further investigated. In addition, some constraints such as size, growing direction and shape could be incorporated into the region growing measure to regularise the resulting regions. The obtained breast segmentation results will be used for further analysis of mammographic images in subsequent chapters.

3.3 A Modified FCM Algorithm for Breast Density Segmentation

The fuzzy c-means (FCM) algorithm has been adopted in a variety of medical image segmentation applications (Sun *et al.*, 2004; Hassanien, 2007). In mammographic image analysis, FCM has been applied to breast density segmentation (Oliver *et al.*, 2008; Lao & Huo, 2009; Keller *et al.*, 2011, 2012). When using FCM for image segmentation, pixels can belong to multiple classes with varying degrees of membership. This has benefits when compared to hard segmentation methods (Pham & Prince, 1998). The conventional FCM algorithm uses greylevel information at a single pixel as the feature space and this contains no spatial contextual information, which makes it very sensitive to noise and intensity inhomogeneity (Chuang *et al.*, 2006). Recently, some modified FCM algorithms with spatial constraints have been published (Ahmed *et al.*, 2002; Szilagyi *et al.*, 2003; Cai *et al.*, 2007; Szilágyi *et al.*, 2007; Kang *et al.*, 2009). However, these algorithms have individual disadvantages and are not robust with respect to varieties of noise. In this section, we present a modified FCM algorithm incorporating local spatial and intensity information based on an adaptive local window filter. The clustering procedure is performed on the intensity histogram of the filtered image rather than every single pixel, which can drastically reduce the computational cost of large-sized mammographic images.

3.3.1 FCM Algorithms

The standard FCM algorithm partitions pixels into categories using a fuzzy membership function (Bezdek, 1981). Let $X = (x_1, x_2, \ldots, x_N)$ denote an image with N pixels to be partitioned into c clusters. The algorithm is an iterative optimisation

that minimises the objective function defined as:

$$J_{FCM} = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \|x_k - v_i\|^2, \tag{3.1}$$

with the following constraints: $\sum_{i=1}^{c} u_{ik} = 1$ for $\forall k$ and $0 < \sum_{k=1}^{N} u_{ik} < N$ for $\forall i$, where u_{ik} represents the membership value of pixel k to cluster i, x_k represents the intensity value of pixel k, and v_i is the prototype value of the i^{th} cluster centre. The parameter m (equal to 2 in our work) is a weighting exponent on each fuzzy membership that controls the fuzziness of the resulting partition.

3.3.1.1 Bias-Corrected FCM (BCFCM) Algorithm

Ahmed *et al.* (2002) modified the objective function of the standard FCM algorithm to compensate for intensity inhomogeneity. The labelling of a pixel takes the labels in its immediate neighbourhood into account. This regularisation results in piecewise-homogeneous labelling. The objective function of BCFCM is given by:

$$J_{BCFCM} = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \left(\|x_k - v_i\|^2 + \frac{\alpha}{|N_k|} \sum_{r \in N_k} \|x_r - v_i\|^2 \right), \tag{3.2}$$

where N_k denotes the neighbouring pixels within a window around pixel k, and $|N_k|$ is the cardinality of N_k. The effect of the neighbour term is controlled by the parameter α. This algorithm can produce better results than FCM in segmenting scans corrupted by salt and pepper noise (also known as shot or impulse noise).

3.3.1.2 Enhanced FCM (EnFCM) Algorithm

Szilagyi *et al.* (2003) proposed the Enhanced FCM (EnFCM) algorithm to accelerate FCM and BCFCM by grouping the pixels possessing the same intensity value in the iterative clustering. In this algorithm, an averaging filter is applied first to compute a linearly-weighted sum image which is described as:

$$\xi_k = \frac{1}{1+\alpha} \left(x_k + \frac{\alpha}{|N_k|} \sum_{r \in N_k} x_r \right) = \frac{1}{1+\alpha} \left(x_k + \alpha \overline{x}_k \right). \tag{3.3}$$

The objective function of the EnFCM algorithm corresponding to the generated sum image ξ is defined as:

$$J_{EnFCM} = \sum_{i=1}^{c} \sum_{l=1}^{q} h_l u_{il}^{m} \|\xi_l - v_i\|^2, \tag{3.4}$$

where h_l stands for the number of pixels in the image ξ with a greylevel value equal to l, ξ_l denotes the intensity value of the l^{th} greylevel, and q is the number of greylevels in ξ. Because the number of greylevels is generally much smaller than the number of pixels in an image, the EnFCM algorithm can drastically reduce the computational complexity of the FCM and BCFCM algorithms.

3.3.1.3 Local Spatial and Intensity Information

To eliminate the aspects associated with the parameter α in Equations 3.2 and 3.3, Cai *et al.* (2007) proposed the fast generalised fuzzy c-means (FGFCM) algorithm. In this algorithm, a factor $S_{ij} = S_{s_{ij}} \cdot S_{g_{ij}}$ is defined as a local similarity measure to replace α, which can provide robustness to noise and detail-preserving with respect to the original image. Here i represents the central pixel of a local window and j represents the neighbouring pixels of i within the local window. $S_{s_{ij}}$ is based on an exponential function with the maximum x or y distance between i and j as an argument, while $S_{g_{ij}}$ is also based on an exponential function using the intensity difference as an argument (and assuming a Gaussian intensity distribution).
Subsequently, Szilágyi *et al.* (2007) suggested a variation on the approach of Cai *et al.* (2007). For $S_{s_{ij}}$ an exponential function is used with a Euclidean distance between pixels i and j as an argument, while for $S_{g_{ij}}$ a cosine based function is used. For the neighbouring pixels whose intensity values are significantly different when compared to the central pixel, the value of $S_{g_{ij}}$ is set to zero. They indicated improved performance with respect to their earlier work (Szilagyi *et al.*, 2003).
Another closely related work was proposed by Kang *et al.* (2009). They introduced an adaptive weighted averaging filter. In this filter, a non-linear weighting is used. However, the resulting weightings are binary and equal to a constant value $k_0 \in < 0, 0.5 >$ when the intensity difference with respect to the central pixel is above a threshold (k_1), or equal to 1 otherwise.
The above FCM based algorithms have their individual disadvantages. An apparent disadvantage of the standard FCM algorithm is that it takes no account of spatial information, which increases its sensitivity to noise and intensity inhomogeneity. One disadvantage of the BCFCM algorithm is that it computes the neighbour term in every iteration step, which significantly increases the computational complexity. Another disadvantage of BCFCM is that the parameter α is chosen experimentally, which makes it difficult to control the effect of the neighbour term. This is also the disadvantage of the EnFCM algorithm which involves the parameter α when computing the resulting sum image with the averaging filter. In addition, one common disadvantage of the approaches of Cai *et al.* (2007) and Szilágyi *et al.* (2007) is that when computing the similarity mea-

sure with respect to the central pixel, most (if not all) the neighbouring pixels are considered as reliable and included to calculate the resulting intensity value of the central pixel. This seems not to fully take into account the fact that large intensity differences between pixels within a local window indicate a high probability of the existence of noise or pixels belonging to different clusters within the current neighbourhood. This shortcoming still indicates a lack of robustness with respect to noise and outliers, especially impulse and mixed noise. The noise and outliers therefore should be detected as unreliable and excluded from estimating the similarity measure.

3.3.2 A Modified FCM Algorithm

Motivated by the advantages and disadvantages of the range of FCM based algorithms discussed above, we present a set of modifications to the EnFCM and FGFCM algorithms and propose a modified FCM (MFCM) algorithm incorporating local spatial and greylevel information. We define an adaptive local window filter which differentiates the weight of the neighbouring pixels when computing the filter response value of the central pixel. The weighting coefficients of the neighbouring pixels are automatically determined based on their spatial and greylevel associations with the central pixel. The filtering is a two-pass process: the first pass is evaluating the neighbouring pixels to distinguish between unreliable and reliable neighbours; the second pass is computing the new intensity value of the central pixel using only the reliable neighbours to generate the filtered image. Finally, a fast clustering is performed based on the intensity histogram of the filtered image. The MFCM algorithm is described as follows:

1. Define a local square window N_k centred on pixel k. The window size used throughout this work is 5×5 (other shapes and sizes are possible).

2. Find the median intensity value $\overline{x}_k$ and compute the deviation σ_k from $\overline{x}_k$ within N_k.

$$\sigma_k = \sqrt{\frac{1}{n_k} \sum_{r \in N_k} (x_r - \overline{x}_k)^2}, \tag{3.5}$$

   where x_r denotes the intensity value of pixel r within N_k, and n_k (equal to 25 in this work) is the number of pixels within N_k. If the difference between x_r and $\overline{x}_k$ is larger than σ_k, pixel r is evaluated as unreliable, otherwise as reliable. We use N_r to represent the reliable neighbouring set, n_r ($n_r \leq 24$, as the central pixel k is excluded) is the number of pixels within N_r.

3. Define a local window filter and compute the weighting coefficients C_{kr}.

$$C_{kr} = \begin{cases} C_{kr_s} \cdot C_{kr_g} & \text{if } r \in N_r \\ 0 & \text{otherwise} \end{cases}, \tag{3.6}$$

where C_{kr_s} and C_{kr_g} are the spatial term and the greylevel term, determined by the local spatial distance and the local intensity difference between the neighbouring pixel r and the central pixel k. They are defined as:

$$C_{kr_s} = \begin{cases} \exp\left(\frac{-d_{kr_s}}{\lambda_s}\right) & \text{if } r \in N_r \\ 0 & \text{otherwise} \end{cases}, \tag{3.7}$$

where d_{kr_s} is the Euclidean distance between pixels k and r, λ_s is the scale factor which determines the change characteristic of C_{kr_s}, and

$$C_{kr_g} = \begin{cases} \exp\left(\frac{-(x_k - x_r)^2}{\lambda_g \cdot \sigma_{kr_g}^2}\right) & \text{if } r \in N_r \\ 0 & \text{otherwise} \end{cases}, \tag{3.8}$$

where x_k denotes the intensity value of the central pixel k, λ_g is the effect factor of greylevel information which controls the influence extent of the neighbouring pixels. $\sigma_{kr_g} = \sqrt{\left(\sum_{r \in N_r} (x_r - x_k)^2\right) / n_r}$ is the intensity deviation from the central pixel k within the reliable neighbouring set N_r. C_{kr_g} is defined as described in (Cai *et al.*, 2007).

4. Filter the image using the defined local window filter. The resulting intensity value ξ_k of the central pixel k is computed by:

$$\xi_k = \left\{ \frac{\sum_{r \in N_k} C_{kr} \cdot x_r}{\sum_{r \in N_k} C_{kr}}. \right. \tag{3.9}$$

Two examples of local window filtering are shown in Figure 3.9. The upper number in every window cell is the intensity value of each pixel, while the lower number is the weighting coefficient of the local window filter ($\lambda_s = 1$ and $\lambda_g = 3$). The local window on the left is almost homogeneous, and the original intensity value 110 of the central pixel remains equal to 110 after local window filtering. By contrast, the local window on the right is inhomogeneous, and the original intensity value 110 of the central pixel is replaced by the response value 60 of the local window filter. The circled neighbouring pixels which might be noise (red) or might belong to different clusters (blue or green) are evaluated as unreliable and excluded from computing the filter response (weighting coefficients equal to 0). The two examples indicate that

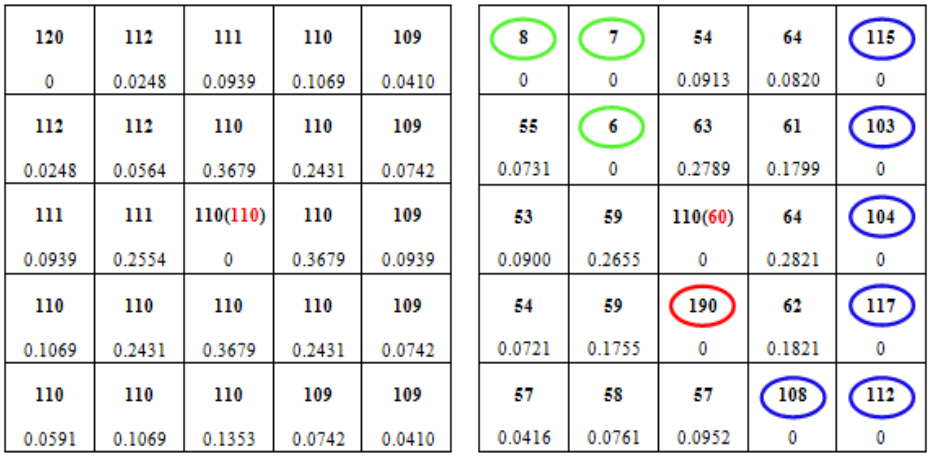


Figure 3.9: Two examples of local window filtering (see main text for details). On the left the response is not altered, but on the right it is.

the defined local window filter can preserve local intensity homogeneity and is robust to noise and intensity inhomogeneity.

5. Count the intensity histogram of the filtered image to obtain the number of greylevels (q in Equation 3.4) and the number of pixels with the same greylevel (h_l in Equation 3.4).

6. Cluster the filtered image based on its intensity histogram using the EnFCM algorithm. The computational complexity of MFCM is $O(qcI)$, where I is the iteration number in the clustering process, which is far lower than $O(NcI)$ of FCM ($q \ll N$).

As described above, the proposed local window filter has the capability to remove noise, exclude outliers, and eliminate unnecessary blur. Moreover, the separation of filtering from clustering can drastically reduce the running time.

3.3.3 Experimental Results

To demonstrate the robustness and effectiveness of the proposed algorithm, it has been tested on both synthetic and mammographic images. The experimental results on synthetic images are quantitatively compared with four algorithms: FCM, Kang *et al.* (2009), Cai *et al.* (2007), and Szilágyi *et al.* (2007), as described in Section 3.3.1.

3.3.3.1 Experimental Results on Synthetic Images

The synthetic images (200×200 pixels) contain four classes (100×100 pixels in each class) corresponding to intensity values 50, 100, 150, and 200, respectively. The

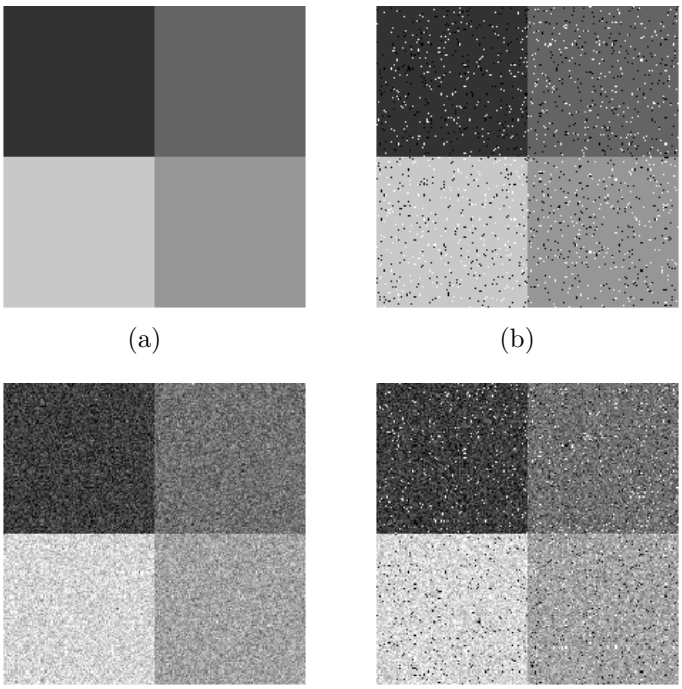


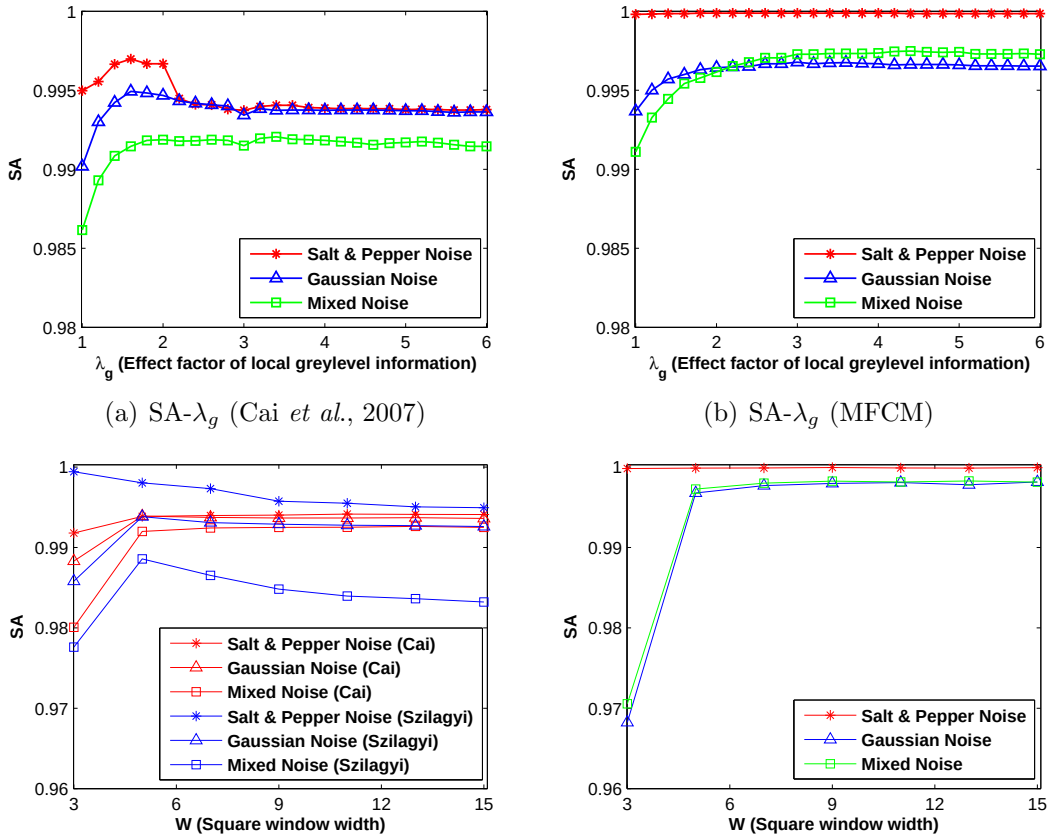
Figure 3.10: Example synthetic images: (a) original synthetic image; (b) add 5% salt and pepper noise; (c) add 5% Gaussian noise; and (d) add 5% mixed noise.

Table 3.4: Means and standard deviations of segmentation accuracy (%) of five algorithms for three types of noise (all noise at 5% level).

Noise Type	FCM	Kang *et al.* (2009)	Cai *et al.* (2007)	Szilágyi *et al.* (2007)	MFCM
Salt & Pepper	96.26±0.10	96.59±0.09	99.39±0.03	99.78±0.03	99.99±0.01
Gaussian	74.93±0.24	99.19±0.05	99.33±0.02	99.35±0.03	99.65±0.03
Mixed Noise	71.36±0.16	96.30±0.09	99.13±0.02	98.79±0.04	99.66±0.04

original images are corrupted by different types of noise: salt and pepper noise, Gaussian noise, and mixed noise (salt and pepper + Gaussian). Figure 3.10 shows an example synthetic image before and after being corrupted by noise. Note that such severe noise (especially for the salt and pepper noise and the mixed noise) rarely occurs in mammographic images. The purpose of corrupting images with a variety of noise at high levels is to test the robustness of the modified FCM algorithm presented in this section, and to make a complete comparison with the other modified FCM algorithms (Kang *et al.*, 2009; Cai *et al.*, 2007; Szilágyi *et al.*, 2007) (the experiments conducted in these papers were all based on synthetic images degraded by the three noise models).

Table 3.4 shows the means and standard deviations of segmentation accuracy (SA) obtained by five algorithms on ten synthetic images for each noise type at the same level (5%). SA is defined as the percentage of correctly classified pixels (with respect to the total number of pixels) (Ahmed *et al.*, 2002). The parameters



(a) SA-λ_g (Cai *et al.*, 2007)

(b) SA-λ_g (MFCM)

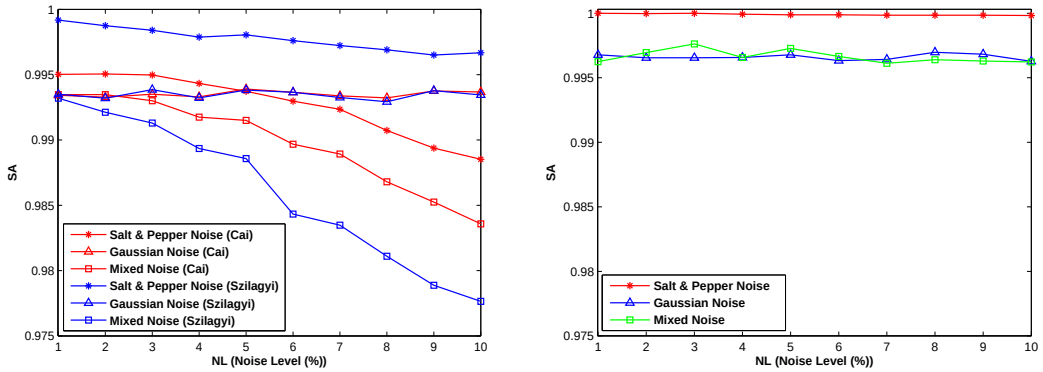
(c) SA-W (Cai *et al.*, 2007, Szilágyi *et al.*, 2007)

(d) SA-W (MFCM)

Figure 3.11: SAs of Cai *et al.* (2007), Szilágyi *et al.* (2007), and MFCM with respect to different values of λ_g and W.

used here are $c = 4$, $m = 2$, $\varepsilon = 0.001$, $k_0 = 0.45$, $k_1 = 0.65$, $\lambda_s = 1$, and $\lambda_g = 3$. The results show that the proposed MFCM algorithm performs better than the other four algorithms for segmenting images corrupted by three types of noise. All algorithms, except Kang *et al.* (2009), obtain their best results on images with salt and pepper noise. For images with mixed noise, all the other four algorithms obtain their worst results, while MFCM still obtains good results, which indicates it performs more robustly with respect to mixed noise.

Figure 3.11 shows the SAs of Cai *et al.* (2007), Szilágyi *et al.* (2007), and MFCM for different values of parameters λ_g and W. The SA of Cai *et al.* (2007) is irregular with respect to the parameter λ_g (Figure 3.11(a)). By contrast, the SA of MFCM is stable for salt and pepper noise and monotonously increases with λ_g and tends to be stable after $\lambda_g = 3$ in the presence of Gaussian and mixed noise (Figure 3.11(b)). Moreover, the SAs of Cai *et al.* (2007) and Szilágyi *et al.* (2007) are not stable for an increasing W (Figure 3.11(c)). By contrast, the SA of MFCM is almost stable when W is larger than 5 (Figure 3.11(d)). Consequently, it is relatively easier to choose appropriate parameters for MFCM. Figure 3.12 shows



(a) SA-NL (Cai *et al.*, 2007, Szilágyi *et al.*, 2007) (b) SA-NL (MFCM)

Figure 3.12: SAs of Cai *et al.* (2007), Szilágyi *et al.* (2007), and MFCM depending on noise levels (%).

the SAs of Cai *et al.* (2007), Szilágyi *et al.* (2007), and MFCM for an increasing noise level. It is shown that the SAs of Cai *et al.* (2007) and Szilágyi *et al.* (2007) decreases as the noise level increases, especially for mixed noise (Figure 3.12(a)). However, the SA of MFCM remains stable with no significant change with respect to increasing noise levels for all the three noise types (Figure 3.12(b)). Both Figures 3.11 and 3.12 indicate the improved overall performance for MFCM when compared to that for Cai *et al.* (2007) and Szilágyi *et al.* (2007).

3.3.3.2 Experimental Results on Mammographic Images

We have applied the MFCM algorithm to mammographic images for breast density segmentation. In this section, the mammographic images from the EPIC database were used for experiments. Twenty-four images were selected covering six different density categories (4 for each SCC density class). The breast region was segmented using the method described in Section 3.2. We down-sampled these images by a factor of 10. The average size of the down-sampled images was 352×214 pixels. These mammographic images were filtered using the same 5×5 local window filter ($\lambda_s = 1$ and $\lambda_g = 3$) as used for the synthetic images. An intensity histogram was generated from the filtered images. The MFCM clustering was performed on this histogram to partition breast tissue pixels into four clusters. After assigning each breast tissue pixel into one corresponding cluster, the breast region was segmented into four uniform density sub-regions: fatty tissue, fatty glandular (semi-fatty) tissue, dense glandular (semi-dense) tissue, and dense tissue. Example segmentation results for six density categories are shown in Figure 3.13. These results indicate reasonable segmentation of the various density regions within the breasts. For the low-density breasts (SCC1 to SCC3), the breast regions are mainly composed of fatty tissue (blue regions) and fatty glandular tissue (light blue regions). On

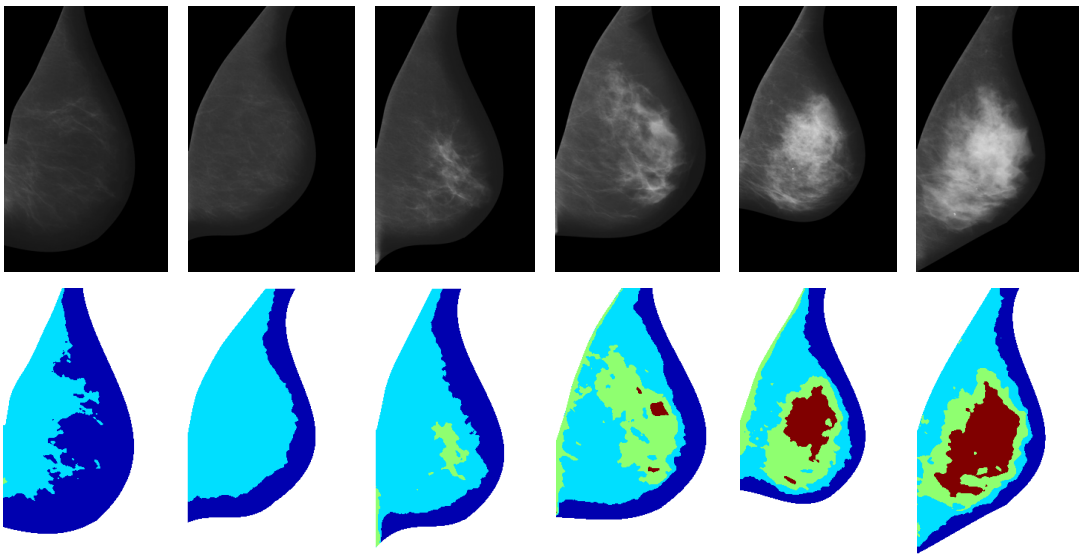


Figure 3.13: Example segmentation results: from left to right, mammograms are sorted from SCC1 to SCC6. Top row: extracted breast regions; bottom row: breast density segmentation results. Different colours represent different density components: fatty tissue (blue), fatty glandular tissue (light blue), dense glandular tissue (green), and dense tissue (red).

the other hand, for the high-density breasts (SCC4 to SCC6), the breast regions contain a relatively high proportion of dense tissue (red regions). The obtained segmentation results can be used for breast density estimation. A further analysis of the MFCM based breast density segmentation for mammographic risk assessment will be provided in Chapter 6.

In addition, we tested the computational time for the standard FCM algorithm and the proposed MFCM algorithm based on the 24 full-resolution mammographic images. As stated in Section 3.3.2, the computational complexity for FCM mainly depends on the number of pixels in the image, while the computational complexity for MFCM is determined by the number of greylevels in the image. The average size of the 24 full-resolution mammographic images was 3516×2136 pixels and the average number of greylevels over the 24 images was 166. Due to the significant difference between the number of pixels and the number of greylevels, the computational time can be dramatically reduced by MFCM. The average computational time was around 1654 seconds and 4 seconds for FCM and MFCM, respectively.

3.4 Topographic Representation Based Breast Density Segmentation

There are a variety of approaches to breast density segmentation in the literature (see Section 2.1.2). A number of previous methods need a learning process

to model mammographic tissue (Petroudi & Brady, 2006; Oliver *et al.*, 2008; He *et al.*, 2011). The iterative optimisation in the procedure of learning can be time consuming. Moreover, some segmentation methods are based on pixel-wise classification (Lao & Huo, 2009; Kallenberg *et al.*, 2011), involving high-dimensional features, which can also result in a high computational cost.

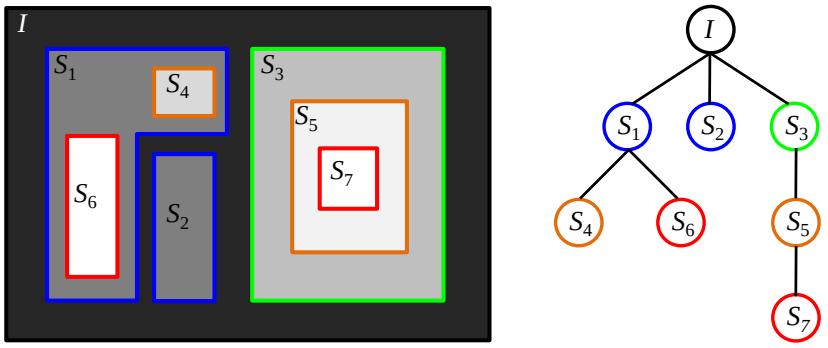
In this section, we present a novel method for breast density segmentation based on topographic representation of mammographic images. The global structure of dense tissue within the breast is analysed based on a topographic map of the whole breast, which is a hierarchical representation, obtained from the upper level sets of the image. A *shape* tree is constructed to represent the topological and geometrical structure of the topographic map. We analyse the *prominency* and *independency* for the *shapes* within the topographic map to detect the dense tissue regions. This method needs neither a learning stage nor prior information about the regions of interest. The hierarchical nature of the topographic map enables the segmentation to be performed at multiple scales. In addition, the analysis based on the components (i.e. *shapes*) of the topographic map instead of the original image pixels can drastically reduce the dimensionality of the data to be analysed. Furthermore, in contrast to segmenting dense tissue by thresholding and thus only using the proportion of dense tissue for mammographic risk assesssment, the density variation within the dense tissue area is also considered by incorporating intensity information into the density map. Closely related work was presented in (Hong & Brady, 2003; Hong & Sohn, 2010), where the topographic approach was used to detect the breast boundary, the pectoral muscle and candidate masses instead of dense tissue regions.

3.4.1 Topographic Representation

We consider an image as a surface where the intensity value at each pixel corresponds to the height. This leads us to build a topographic representation for mammographic images based on intensity information. In mammographic images, the dense tissue regions are high-intensity regions, thus we use the upper level sets to generate a topographic map, which can be naturally represented by a tree due to its hierarchical structure.

3.4.1.1 Topographic Map

A topographic map is a morphological and multiscale decomposition of an image relying on the connected components of the level sets, which can represent both topological and geometrical structure of objects in images. An upper level set of an image $I : \Omega \subset \mathbb{N}^2$ is defined as $U_l(I) = \{(x, y) \in \Omega \mid I(x, y) \geq l\}$, where Ω is



(a) (b)

Figure 3.14: The topographic map and its corresponding *shape* tree of a synthetic image: (a) topographic map; (b) *shape* tree.

the domain of the image I, and $l \in \mathbb{N}$ is a given intensity level. The topographic map is built by the upper level sets at a sequence of intensity levels over the full intensity range of the image. The topographic map of the image I is given by $TM(I) = \{U_{l_i}(I) \mid l_i \in L, i = 1, 2, \cdots, N\}$, where L denotes the intensity range of the image I, and N is the number of intensity levels for generating the upper level sets. We refer to the basic elements of the topographic map as *shapes*, obtained by the connected components (8-connectivity) of the upper level sets in which the holes have been filled (see (Xia *et al.*, 2010) for details). An example topographic map of a synthetic image is shown in Figure 3.14(a). The greylevel value of each pixel in the topographic map is the intensity level at which the smallest *shape* containing the pixel is extracted. The colour contours superimposed on the topographic map are the corresponding level lines.

The upper level sets can be regarded as a complete decomposition of the image, which constitute a decreasing family corresponding to an increasing intensity level. Thus, the topographic map has a multiscale structure from large to small scales. Here, the scale corresponds to the area of the *shapes* in the topographic map. In addition, the number of intensity levels N can be used as a resolution factor. A large N can capture a high-resolution topographic map, while a small N can form a low-resolution topographic map. Furthermore, the topographic map is contrast invariant to any increasing contrast change due to the natural property of the upper level sets.

3.4.1.2 Shape Tree

The global pattern of the *shapes* in the topographic map can be represented in the form of a *shape* tree. The *shape* tree is constructed based on an inclusion relationship between the *shapes*, where each node corresponds to a *shape* and the root node represents the whole image. The inclusion relationship between two *shapes*

can be either internal or external. If *shape* S_j is spatially enclosed by *shape* S_i, S_i is the *parent* node of S_j, and S_j is the *child* node of S_i. The *shape* tree corresponding to the topographic map shown in Figure 3.14(a) is illustrated by Figure 3.14(b). Here, we define some terms based on the *shape* tree in Figure 3.14(b) for further analysis. A node's *degree* is defined as the number of its immediate children nodes. If the *degree* of one node is larger than 1, this node is called a *branching* node. The immediate *child* node of a *branching* node is called a *base* node. The nodes with *degree* equal to 0 are called *terminal* nodes. An M-generation *ancestor family* of a *shape* S, denoted by $\mathcal{A}^M(S)$, is defined as $\mathcal{A}^M(S) = \{P^m(S), \mid m = 1, 2, \cdots, M\}$, where $P^m(S)$ is the m^{th} *parent* of S. An M-generation *descendent family* of a *shape* S, denoted by $\mathcal{D}^M(S)$, is defined as $\mathcal{D}^M(S) = \{C^m(S), \mid m = 1, 2, \cdots, M\}$, where $C^m(S)$ is the m^{th} *child* of S.

3.4.2 Segmentation of Dense Tissue Regions

As a pre-processing step, the breast region is segmented using the method described in Section 3.2 (Figure 3.15(a)). The topographic map superimposed with the level lines (*shape* contours) is shown in Figure 3.15(b) (32 uniformly spaced intensity levels are used and the *shapes* smaller than 300 pixels are removed to eliminate sensitivity to noise and small intensity irregularities). It is shown that the topographic map effectively characterises mammographic densities and extracts regions at a range of densities. In addition, the *shape* contours closely capture the boundaries of regions with various densities. On the one hand, a nesting pattern is formed near the boundary of a distinctively dense region where the *shape* contours are densely nested. On the other hand, a significant transition happens between the contours of two adjacent *shapes* if the smaller one is the outermost boundary of a dense region. As such, dense tissue regions are regarded as *prominent* or *independent shapes* in the topographic map. Specifically, a *prominent shape* is a *shape* which indicates the local maximum intensity value with respect to the surrounding tissue; and an *independent shape* corresponds to a *shape* of which the contour has a significant transition from that of the *parent shape*. Therefore, two properties are defined for the *shapes* to detect dense tissue regions from the topographic map, called *prominency* and *independency*, respectively.

3.4.2.1 *Prominency*

Dense tissue regions are considered as *prominent shapes*, since such regions have higher intensity values than the surroundings. The *prominency* of a *shape* can be evaluated based on the *shape* tree. In each branch of the tree, the nodes closer to the *terminal* node have a higher *prominency* score compared to those

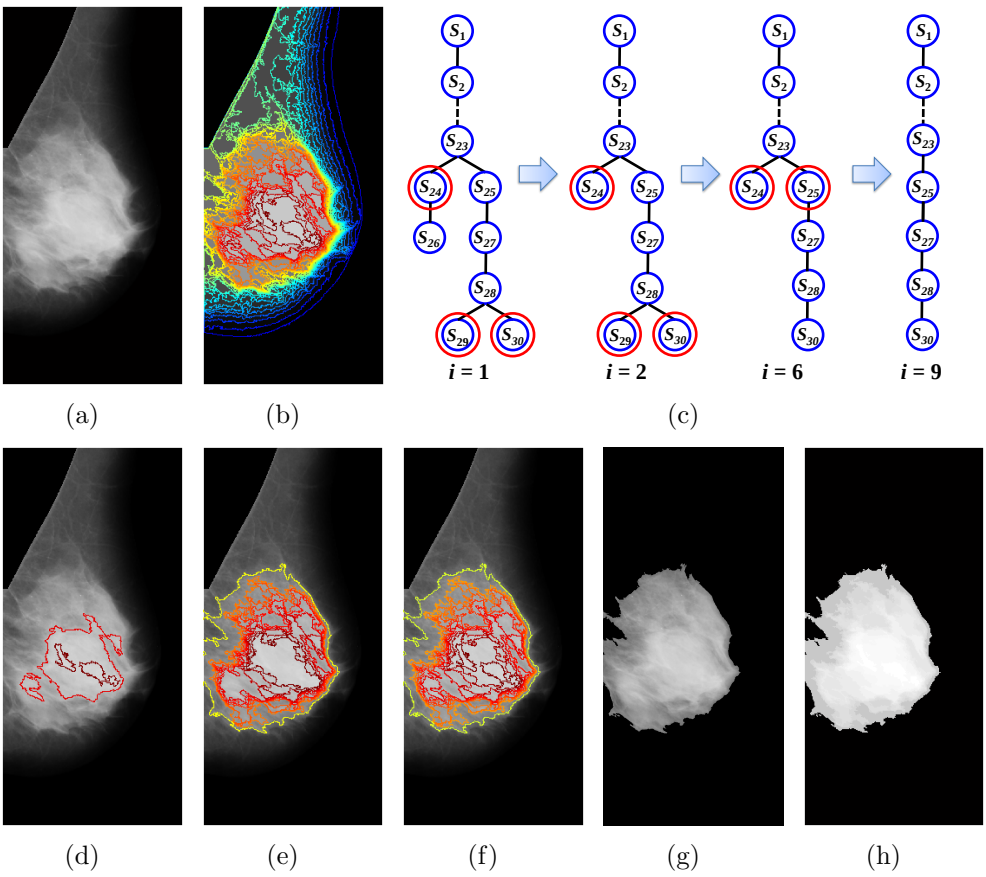


Figure 3.15: Segmentation of dense tissue regions: (a) breast region segmentation; (b) topographic map superimposed with *shape* contours; (c) multiscale *shape* trees (circled nodes are the detected *prominent shapes* at that scale); (d) dense tissue regions extracted based on *prominency*; (e) dense tissue regions extracted based on *independency*; (f) all resulting dense tissue regions; (g) extracted dense tissue; and (h) density map (the contrast has been normalised for better visualisation).

further away, as the *terminal* node indicates the maximum intensity value among all the nodes in this branch. Therefore, the *terminal* nodes are rated as the most *prominent shapes*. However, the *base* node corresponding to each *terminal* node is finally selected, as it is the outermost *shape* enclosing a dense region. Dense tissue regions are detected by iteratively tracing the *base* nodes from a series of *shape* trees from the finest scale to coarser scales, which are generated by gradually removing smaller *shapes* from the *shape* tree (the scale of the *shape* tree indicates the area of the smallest *shape* retained in the *shape* tree). This process terminates when a trunk is formed. Figure 3.15(c) concisely illustrates the truncation process covering the core *shape* trees (*shape* $S_3, S_4, \cdots, S_{22}$ are not displayed as these do not add any information). It is shown that the tree is truncated at scales 2, 6 and 9, and the trunk is formed at the 9^{th} scale. Finally, S_{24}, S_{25}, S_{29} and S_{30} are selected as dense tissue regions shown in Fig. 3.15(d). An algorithmic description of the process of detecting *prominent shapes* is provided in Algorithm 1.

Algorithm 1 The detection of *prominent shapes* by iteratively finding *base* nodes from a series of *shape* trees from the finest scale to coarser scales.

1. Initialise the set of *prominent shapes* to NULL (S_P = NULL).
2. $i = 1$; (Start from the finest scale where the area of the *shapes* is no smaller than 300 pixels).
3. Find all the *terminal* nodes (N_T) of the *shape* tree and compute the number of the *terminal* nodes ($|N_T|$).
4. Trace backwards from each *terminal* node and find the *base* node (N_B) of the tree branch which it belongs to.
5. $S_P = [S_P, N_B]$; (Add all the *base* nodes found at Step 4 to S_P).

while $|N_T| > 1$ (The *shape* tree has more than one branches.) **do**
$i = i + 1$; (Proceed to the next scale.)
Remove the *shapes* smaller than $300 \times i$ pixels from the *shape* tree generated at the previous scale.
Perform Steps 3-5.
end while($|N_T| = 1$; A tree trunk is formed.)

3.4.2.2 *Independency*

Dense tissue regions are regarded as *independent shapes* since a sharp shape transition happens between their *ancestor* and *descendent families*. The *independency* of a *shape* is defined to measure the shape difference between its *ancestor family* and *descendent family*:

$$Independency(S) = \frac{1}{Area(S)} \cdot \frac{\langle Area(\mathcal{A}^M(S))\rangle - Area(S)}{Area(S) - \langle Area(\mathcal{D}^M(S))\rangle}, \tag{3.10}$$

where $Area(S)$ is the area of the *shape* S equal to the number of pixels contained in S, $\langle\cdot\rangle$ is the mean operation, $\langle Area(\mathcal{A}^M(S))\rangle$ and $\langle Area(\mathcal{D}^M(S))\rangle$ are the mean areas of the *ancestor* and *descendent families* of the *shape* S, which are computed by $\frac{1}{M}\sum_{m=1}^{M} Area(P^m(S))$ and $\frac{1}{M}\sum_{m=1}^{M} Area(C^m(S))$, respectively. $Ind(S)$ is normalised by $Area(S)$ to avoid bias caused by the size of the *shape*. It is noted that a large *independency* value indicates the *shape* has significant shape change with respect to its *ancestor family* but insignificant shape change with respect to its *descendent family*, which indicates a high probability of it being the outermost *shape* enclosing a dense region. Thus, dense tissue regions are detected from the resulting trunk of the *shape* tree by setting a threshold on $Ind(S)$. The *shapes* with *independency* values larger than the threshold are selected as dense tissue regions. For an illustration of the detection of subtle region boundaries, we set $M = 1$ and used a threshold of 8.6×10^{-6}. This threshold was determined based on the distribution of the *independency* values of the *shapes* in the resulting trunk computed by Equation 3.10. Specifically, these *independency* values were first sorted in descending order and the gradient of the sorted values was computed

within a neighborhood of three points. Then the first point of zero gradient was found and its *independency* value was chosen as the threshold. This process can be regarded as finding the point of inflexion of the sorted *independency* values of the *shapes* in the trunk. As a result, S_{15}, S_{19}, S_{20}, S_{21}, S_{22}, S_{23}, S_{25}, S_{27} and S_{28} are detected as the *independent shapes*. The corresponding dense tissue regions are shown in Figure 3.15(e).

To remove false positive regions (non-dense tissue regions), two descriptors of each resulting *shape*, *elongation* and *compactness*, are computed by:

$$Elongation(S) = R_{major}(S)/R_{minor}(S), \tag{3.11}$$

where $R_{major}(S)$ and $R_{minor}(S)$ represent the radius of the major axis and the minor axis of the equivalent ellipse of the shape S, respectively.

$$Compactness(S) = 4\pi Area(S)/Perimeter(S)^2, \tag{3.12}$$

where $Area(S)$ and $Perimeter(S)$ represent the area and the perimeter of the shape S, respectively. In our work, the regions of interest are compact, approximately circular or elliptical regions, and therefore the elongated regions with a large *elongation* value (> 5.5) and a small *compactness* value (< 0.04) are discarded (the two thresholds were chosen based on a set of manually selected false positive regions). The final detection result of the dense tissue regions including the *prominent* and *independent shapes* is shown in Figure 3.15(f). The resulting dense tissue is shown in Figure 3.15(g).

3.4.3 Breast Density Quantification

In order to use the segmentation results for breast density quantification, a density map is created for mammographic images based on the segmented dense tissue regions (see Figure 3.15(h) for an example), which is defined as:

$$DM(x,y) = \begin{cases} 0 & \text{if } (x,y) \notin Regions, \\ \frac{1}{|R_{min}|} \sum\limits_{(x,y)\in R_{min}} I(x,y) & \text{if } (x,y) \in Regions, \end{cases} \tag{3.13}$$

where *Regions* denotes the union of the segmented dense tissue regions, R_{min} is the smallest region containing pixel (x,y) in *Regions*, and $|R_{min}|$ is the number of pixels within R_{min}. Two features are calculated based on the density map for breast density quantification, called *density area* and *average density*, respectively. For a mammographic image, the *density area* is defined as the number of the non-zero pixels in the density map, computed by $density\ area = |Regions|$. The

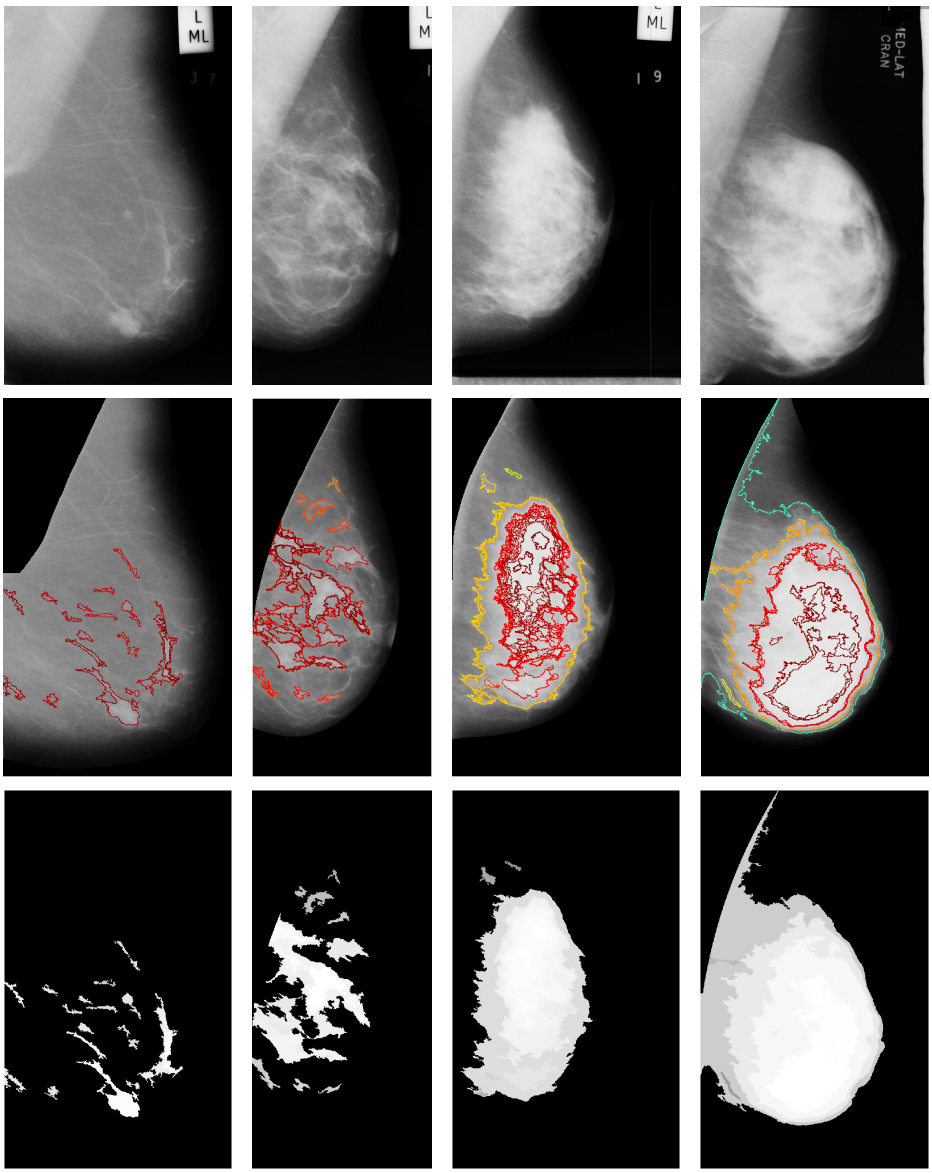


Figure 3.16: Segmentation results and density maps (contrast has been normalised for better visualisation) of example mammograms from the MIAS database. From left to right, the mammograms are sorted from BIRADS I to BIRADS IV.

average density is defined as the average pixel value of the non-zero pixels in the density map, computed by $average\ density = \frac{1}{|Regions|} \sum_{(x,y) \in Regions} DM(x,y)$. The product of the *density area* and *average density* is used as a measure of breast density (denoted by *density*).

3.4.4 Results

To evaluate the proposed method for breast density segmentation and investigate its potential for breast density quantification, it has been tested using two datasets. One is the full MIAS database containing left and right MLO mammograms from

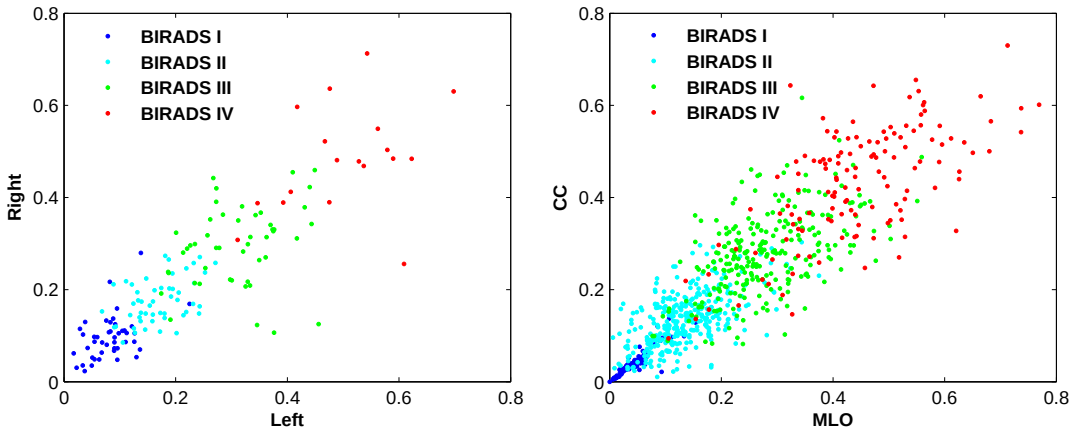


Figure 3.17: Scatter plots of the *density* value of bilateral mammograms and MLO/CC mammograms for the MIAS and DDSM databases.

161 women, the other is taken from the DDSM database containing right MLO and CC mammograms from 831 women.

For each mammogram, a relatively large number of intensity levels ($N = 85$) were used to produce a high-resolution topographic map. As described in Section 3.4.2, the *shape* tree was first generated with the finest scale and it was truncated by gradually removing smaller *shapes* (the area of the *shapes* retained at the i^{th} scale is no smaller than $300 \times i$ pixels). Here, the value of 300 determines the minimum area of the *shapes* contained in the topographic map as well as the step size when truncating the *shape* tree, which was set empirically according to the spatial resolution of mammographic images used for experiments ($200\mu m \times 200\mu m$ per pixel). A setting smaller than 300 can increase the structural complexity of the *shape* tree and decrease the speed of the formation of the tree trunk. For the purpose of the segmentation of dense tissue regions, the setting of 300 is deemed acceptable. A three-generation family ($M = 3$) was used to detect the *independent shapes* from the trunk. Figure 3.16 shows the segmentation results of the dense tissue regions and the corresponding density maps of example mammograms from the MIAS database, covering the four BIRADS density categories. Reasonable segmentations are indicated with respect to the various density regions within the breasts. For BIRADS I, the breast is almost entirely fatty, only a few small regions are segmented; for BIRADS II, some scattered medium regions are segmented as some fibroglandular tissue; for BIRADS III, the breast is heterogeneously dense, larger regions are segmented which include a few small homogeneous regions; while for BIRADS IV, the breast is extremely dense, nearly the whole breast is regarded as the dense tissue region, and large homogeneous regions are segmented.

For breast density quantification, the breast density measure (i.e. *density*) of each

mammogram was calculated as described in Section 3.4.3. The *density area* was normalised by the area of the individual breast to avoid bias caused by the breast size. The *average density* was normalised by a uniform intensity value of 255 to be consistent with the original intensity correlation between mammograms. In addition, we investigated variations of the defined breast density measure between the left and right mammograms of the same woman, and its variations between the MLO view and the CC view mammograms of the same breast. Strong correlations between the left and right breasts and between the two standard views of each breast have been reported in the previous studies (Byng *et al.*, 1996b; Lao & Huo, 2009). Figure 3.17 shows the scatter plots of the *density* value of 160 pairs of left and right mammograms from the MIAS database (mdb295ll and mdb296rl were excluded for historical reasons, and the BIRADS label of the left mammogram was used to determine the colour of the point), and 831 MLO mammograms and 831 CC mammograms from the DDSM database. A strong positive correlation between both sides (Pearson $r = 0.85$, Spearman $rho = 0.85$) and both views (Pearson $r = 0.88$, Spearman $rho = 0.89$) is indicated, which is larger than Pearson $r = 0.78$ in (Lao & Huo, 2009).

3.5 Summary

In this chapter, we presented a review of image segmentation methods and our methods for mammographic image segmentation. We briefly reviewed different categories of image segmentation methods and discussed their core ideas and main advantages and disadvantages.

The developed breast region segmentation method is based on a combination of some reviewed segmentation methods which include histogram thresholding, edge detection in scale space, active contour model, and region growing. Initial segmentation results for the EPIC and MIAS databases have been evaluated for the purpose of further processing in CAD schemes. A high percentage of *acceptable* segmentation results have been obtained for both the breast-background segmentation (98.4% for EPIC and 100% for MIAS) and the pectoral muscle segmentation (93.5% for EPIC and 97.8% for MIAS).

In addition, two breast density segmentation methods have been proposed. The first breast density segmentation method uses the MFCM algorithm to segment the breast region into a number of sub-regions corresponding to different densities. The MFCM algorithm incorporates local information into the standard FCM algorithm by the defined local window filter, which uses the spatial distance and the intensity difference between the neighbouring pixel and the central pixel as the arguments to compute the weighting coefficients. The validity of the MFCM algo-

rithm has been tested using synthetic images corrupted by different types of noise within a range of percent levels. The higher segmentation accuracies have been obtained by MFCM compared to recent algorithms, which indicates its robustness and effectiveness to different noise types and levels. For segmenting large-sized mammographic images, the higher computational efficiency has been indicated compared to the standard FCM.

The second breast density segmentation method uses the topographic representation composed of *shapes* obtained at a sequence of intensity levels to analyse the global structure of dense tissue. The *prominency* and *independency* of the *shapes* were defined to detect the *prominent* and *independent* regions as the dense tissue regions. The intensity variations within the segmented dense tissue besides the area were considered when computing the breast density measure based on the defined density map. A strong positive correlation for both sides (left and right) and both views (MLO and CC) has been indicated with respect to the resulting density measure. Both segmentation methods have indicated reasonable segmentations with respect to breast tissue density. Further evaluation for mammographic risk assessment will be provided in Chapter 6.

Chapter 4

Texture Analysis

Texture analysis is an important area of study in computer vision. Approaches are still being developed for texture modelling, segmentation and classification. There is no precise definition of texture due to its wide variability (Sonka *et al.*, 2007). Intuitively, textures are characteristic intensity (or colour) variations that are typically formed by the surface or internal structure of an object in reflective or transmissive images (Sonka *et al.*, 2007; Mirmehdi *et al.*, 2009). Texture analysis techniques can be applied to many studies of medical images, such as anatomical structure segmentation, lesion detection, healthy and pathological tissue characterisation, and classification (Castellano *et al.*, 2004). For mammographic image analysis, texture features have been used for characterising mammographic parenchymal patterns (see Section 2.3). In this chapter, a brief review of texture analysis approaches is first provided, which focuses on the various methods for texture feature extraction. Subsequently, a number of local feature based texture representations are investigated. On the basis of this, an approach to modelling mammographic tissue appearance based on local tissue appearance is presented.

4.1 A Brief Review of Texture Analysis

In texture analysis, there are such a great diversity of approaches that it is impossible to review all of them thoroughly. In this section, we focus on the widely used methods for extracting texture features and outline these methods into four categories: statistical methods, structural methods, signal processing methods, and model based methods.

4.1.1 Statistical Methods

The spatial distribution of pixel values in an image is one of the defining qualities of texture and plays an important role in describing texture. Using statistical

features to measure the spatial distribution of intensity values is one of the early methods in the computer vision literature (Tuceryan & Jain, 1993). A large number of statistical texture features have been proposed, covering various statistical texture measures (Haralick, 1979). A most commonly used statistical description of intensity values is the intensity histogram, which is the first-order statistical analysis of the image. A number of features can be computed using the intensity histogram, such as mean, variance, skewness, kurtosis, energy and entropy. Some texture features are based on the second-order statistics of image properties, where the relationship between a pair of pixels is analysed to describe the spatial configuration of texture. Haralick *et al.* (1973) proposed to use a co-occurrence matrix to describe two-dimensional spatial dependence of greylevels for a fixed distance and/or angular spatial relationship. A set of texture features can be extracted from this matrix to measure texture characteristics within the image, such as homogeneity, contrast and correlation. This method has become one of the most widely used texture features. A variation on the co-occurrence matrix is based on a greylevel difference method (Conners & Harlow, 1980), which describes the distribution of pixel pairs separated by a fixed distance and having a fixed greylevel difference. Various features can be derived from this matrix for texture analysis, such as contrast, angular second moment, entropy, and inverse difference moment. In contrast, Galloway (1975) proposed a greylevel run length method, which is based on computing the number of greylevel runs of various lengths. A gray level run is a set of linearly adjacent pixels having the same greylevel value. The length of the run is the number of pixels within the run.

In addition, an important property of many textures is that they are composed of repetitive texture primitives. The autocorrelation function of an image can be used to estimate the spatial size of the texture primitives and the fineness/coarseness of the texture: coarse textures are generated from larger primitives, while fine textures are built from smaller primitives. In an autocorrelation texture model, texture can be described using the correlation coefficient which evaluates the linear spatial relationship between primitives (Sonka *et al.*, 2007). In addition, some methods are based on statistical analysis of these fundamental texture elements across the image. A detailed description of a number of texture representations based on local features will be provided in Section 4.2.

4.1.2 Structural Methods

As mentioned above, some textures are composed of texture primitives or elements and are regarded as being generated by a placement of these primitives according to a certain placement rule. Structural methods are based on analysing

the structural properties of these texture primitives in order to extract or generalise the placement rule that characterises the texture (Haralick, 1979). A few representative methods are provided below.

Zucker (1976) proposed aspects of a texture model in which real textures were regarded as distorted versions of ideal textures. The spatial placement rule was defined by a graph that described the spatial relationships between texture primitives. An underlying ideal texture has a regular or isomorphic graph representation and it can be transformed into a realistic texture by distorting the graph. Lu & Fu (1978) developed a syntactic method for generating and discriminating textures. A texture pattern was first divided into small square windows. The spatial structure of unit patterns belonging to the same texture pattern was then represented by a tree grammar. This tree grammar can be used for the synthesis and discrimination of textures. Ahuja (1982) proposed a model for defining the neighbourhood of a point called Voronoi tessellation. The region enclosed by a point's Voronoi polygon was assumed to possess intuitively appealing characteristics and was regarded as the neighbourhood of the point. On the basis of that, Tuceryan & Jain (1990) developed a texture segmentation method in which the local relationships of texture primitives (tokens) were defined using a graph structure. The Voronoi tessellation of the given texture was first built, and then shape features were extracted from the resulting Voronoi polygons.

In addition, Marr (1982) proposed a concept of primal sketch, which is a symbolic representation of an image. Firstly, distinctive image primitives were extracted, such as bars, edges and blobs. Then, the primal sketch is generated from which a set of features can be extracted for texture analysis, such as occurrences of different types of primitives, element orientation, and spatial density of elements. Subsequently, Guo *et al.* (2003) proposed a mathematical model for Marr's primal sketch. They defined a mathematical quantity for sketchability based on the wavelet/sparse coding theory and the Markov random field theory, which can be used to determine if an image portion is sketchable and divide the image into sketchable regions (geometry) and non-sketchable regions (texture).

4.1.3 Signal Processing Methods

Psychological studies have indicated that the human brain does a spatial frequency analysis in visual perception of the image (Bruce *et al.*, 1996). Due to its natural properties, the texture image is well suited to this type of analysis (Tuceryan & Jain, 1993). Thus, many methods for texture analysis are based on signal processing techniques. Most signal processing based methods utilise filters to capture texture properties, and texture features are extracted from filtered images.

Spatial domain filters are the most direct way to extract texture properties from images. Since the density of edges per unit area can characterise the fineness of textures (fine textures generally have a higher density of edges than coarse textures), edge detection filters have been straightforwardly employed to measure edge density, such as Roberts and Laplacian operators (Laws, 1980). Malik & Perona (1990) used a bank of even-symmetric linear filters to model preattentive texture perception in the human visual system. Unser & Eden (1990) used a combination of spatial filters and nonlinear operators to extract texture features. In addition, spatial filters can be used to compute spatial moments of an image region (Laws, 1980), and the resulting moment based features have been applied to texture segmentation (Tuceryan, 1994).

Another set of texture features are extracted from the frequency domain (Tuceryan & Jain, 1993). D'Astous & Jernigan (1984) used measures of the power spectrum to discriminate textures, including peak features such as the strength and area of peaks, and power distribution features such as power spectrum eigenvalues and circularity. On the other hand, since filtering/convolution in the spatial domain is exactly equivalent to multiplication in the frequency domain, the spatial filtering operation can be achieved by performing the multiplication operation in the frequency domain, especially when the filter operator in the spatial domain is difficult to generate. The image is first transformed into the Fourier domain and then multiplied with the filter function in the Fourier domain. After that, the resulting image is inverse-transformed into the spatial domain (Mirmehdi *et al.*, 2009). A number of texture segmentation techniques based on the spatial/spatial-frequency domain analysis are reviewed in (Reed & Wechsler, 1990).

The Fourier transform analyses the global frequency components of the entire texture image, which takes no account of the spatial aspect and has a low spatial resolution (Mirmehdi *et al.*, 2009). Many applications also require local analysis in the spatial domain (Bovik *et al.*, 1990). In order to introduce the spatial dependency into the Fourier analysis, an effective way is using the window Fourier transform (Tuceryan & Jain, 1993). When using Gaussian functions as the window functions, this transform becomes the well-known Gabor transform. Gabor functions are well suited to multiresolution spatial-frequency analysis as they provide an optimal spatial resolution for a given bandwidth (Daugman *et al.*, 1985). Moreover, Gabor functions have been demonstrated to have similar behaviors to the human visual system and to be good fits to the receptive field profiles of simple cells in the striate cortex (Marčelja, 1980). Due to the properties of Gabor functions, Gabor filters have been regarded as a very useful tool in texture analysis. Jain & Farrokhnia (1991) used a bank of Gabor filters to extract texture features for texture segmentation. Idrissa & Acheroy (2002) applied the Gabor filtering

to texture classification, where texture features are obtained by computing local energy using filter response. Rotate invariant texture features based on the Gabor transform are defined in (Arivazhagan *et al.*, 2006) for texture classification. Texture features are derived by calculating the mean and variance of the filtered image. Rotation invariance is achieved by circularly shifting the feature elements, so that all images have the same dominant direction.
In addition, it has been shown that the wavelet transform provides a precise and unifying framework for the analysis and characterisation of a signal at different scales (Rioul & Vetterli, 1991). Due to its capability in multiresolution analysis, the wavelet transform has been used for the characterisation of texture properties at multiple scales (Unser, 1995).

4.1.4 Model Based Methods

Model based texture analysis methods try to construct a texture model which can be used to describe and synthesise textures (Tuceryan & Jain, 1993). In general, the model based methods use stochastic and generative models to represent and generate textures, and the estimated model parameters are used as texture features for texture analysis (Mirmehdi *et al.*, 2009). Markov random fields have been widely used for modelling images, which can capture local contextual information in images (Li, 2009). Markov random field models assume that the intensity value at each pixel in the image only depends on the intensity values of pixels in the local neighbourhood. Cross & Jain (1983) first explored the use of Markov random fields as texture models. Afterwards, numerous approaches based on Markov random field models have been proposed for texture synthesis, texture segmentation and classification. On the other hand, many textures have a statistical quality of roughness and self-similarity at different scales (Tuceryan & Jain, 1993). Fractals have a good capability in modelling these properties of textures. Therefore, the fractal model has received considerable attention in texture analysis (Pentland, 1984), where most features are based on fractal dimension and lacunarity (Keller *et al.*, 1989; Chaudhuri & Sarkar, 1995). However, this model is regarded not appropriate for characterising local image structures (Mirmehdi *et al.*, 2009). In addition, Jojic *et al.* (2003) proposed a novel appearance and shape model called *epitome*. The epitome of an image is a miniature and condensed version of the image, which contains its essential texture and shape properties.

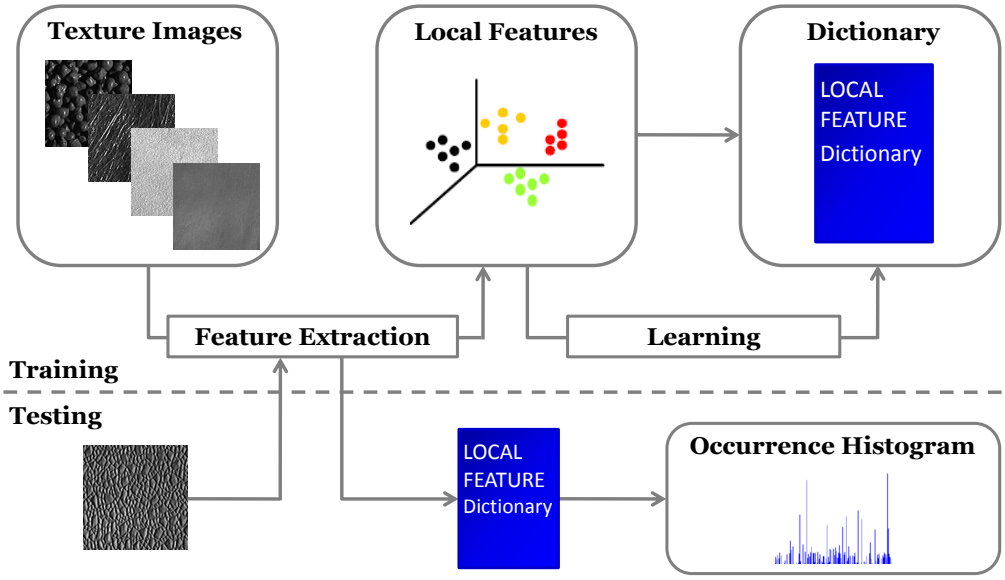


Figure 4.1: The schema of representing textures based on local features.

4.2 Local Feature Based Texture Representations

Recently, local feature based texture representations have shown to be effective for texture classification (Ojala *et al.*, 2002; Varma & Zisserman, 2005; Zhang *et al.*, 2007; Varma & Zisserman, 2009; Crosier & Griffin, 2010). Texture images are statistically analysed by extracting appearance based features from local image patches, and are represented as histograms of a local feature dictionary. The schema of representing textures based on local features is shown in Figure 4.1. Firstly, local features are extracted from local neighbourhoods of image pixels. After that, a visual dictionary of local features is generated. There are two ways of collecting the visual dictionary: one is intuitively aggregating all possible appearances of a specific local feature, which is an exhaustive collection of qualitative categories of local image appearances; the other is performing an initial learning step such as clustering to learn clusters of local features and the cluster centres are collected into the dictionary. After the dictionary of local features is obtained, for a novel texture image, each pixel is labelled by searching for the corresponding local feature ("word") in the dictionary. Finally, the image is represented as an occurrence histogram of the "words" in the dictionary, which provides a statistical description of local feature distribution in the image.

In this section, five different strategies based on local features are investigated: local binary patterns, local greylevel appearances, local geometric structures, joint filter responses of filter banks, and raw image patches.

4.2.1 Local Binary Patterns

Local Binary Patterns (LBP) was first proposed in (Ojala *et al.*, 2002). The local texture information in texture images is encoded into a set of binary values. The greylevel value of the centre pixel is subtracted from the local neighbourhood, and a binary label is assigned to each pixel within the local neighbourhood according to the difference sign. The resulting binary pattern is transformed into a unique LBP number by:

$$LBP = 1 + \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \tag{4.1}$$

where g_c represents the greylevel value of the centre pixel, $g_p(p = 0, \ldots, P-1)$ corresponds to the greylevel value of the p^{th} pixel in the local neighbourhood, and $s(x) = 1$ if $x \geq 0$ else $s(x) = 0$. Thus, each LBP number corresponds to a unique local binary pattern, and all possible binary patterns comprise the visual dictionary of local features. An LBP histogram is populated by counting occurrences of LBP numbers at every pixel to represent the texture image.

4.2.2 Local Greylevel Appearances

The Local Greylevel Appearances (LGA) approach was presented in (Zwiggelaar, 2010). The local texture information is modelled by analysing the joint greylevel distribution of the local neighbourhood. The local greylevel appearance is transformed into a unique LGA number. The distribution of local greylevel appearances in histogram format constructs the basis for texture models. Firstly, the greylevel resolution is reduced to N_g, which is the number of reduced-resolution greylevel bins, and each pixel within the local neighbourhood is numbered in row order starting from zero. Subsequently, the unique LGA number corresponding to the local neighbourhood is computed by:

$$LGA = 1 + \sum_{i,j} N_g{}^{counter(i,j)} I(i,j), \tag{4.2}$$

where $counter(i,j) = 0, \ldots, N-1$ is the sequence number of pixel (i,j), N is the number of pixels in the local neighbourhood, and $I(i,j)$ is the reduced-resolution greylevel value of pixel (i,j). Thus, each LGA number corresponds to a unique local greylevel appearance, and all possible greylevel appearances comprise the visual dictionary. Finally, an LGA histogram containing the combination of LGA numbers and corresponding occurrences is generated.

4.2.3 Basic Image Features

Basic Image Features (BIF) were defined in (Crosier & Griffin, 2010). A second-order family of six Gaussian derivative filters are used to analyse texture images locally with respect to geometric structures. Figure 4.2 shows the Gaussian filter bank at four scales. Seven Basic Image Features are defined, which are referred to as *BIFs*, each corresponding to a qualitatively distinct type of local geometric structures (example image patches are illustrated in Figure 4.3). The seven *BIFs* are respectively defined as: $\left\{\varepsilon s_{00}, 2\sqrt{s_{10}^2 + s_{01}^2}, \pm\lambda, 2^{-\frac{1}{2}}(\gamma \pm \lambda), \gamma\right\}$, where ε is a constant which controls the extent of an image patch to be classified as the flat category (the first one of the seven geometric categories in Figure 4.3). In order to calculate *BIFs*, the filter responses are scale-normalised by $s_{ij} = \sigma^{i+j} c_{ij}$, where s_{ij} is the normalised value of the filter response c_{ij} at scale σ. γ and λ are computed by $\lambda = s_{20} + s_{02}$ and $\gamma = \sqrt{(s_{20} - s_{02})^2 + 4s_{11}^2}$, respectively. The filter response space is divided into seven regions, and each corresponds to a specific geometric structure. The geometric structure of a local image patch is determined according to the largest *BIF* computed using the filter responses of its centre pixel. Thus, a texture image can be modelled as a histogram over the seven *BIFs*. This modelling procedure of populating a 7-bin *BIF* histogram can be extended to generate a more descriptive texture representation by analysing the geometric structures of local image patches at multiple scales. The whole modelling procedure is: 1) convolve the image with the Gaussian filter bank at four octave-separated scales (σ_{base}, $2\sigma_{base}$, $4\sigma_{base}$, $8\sigma_{base}$); 2) compute *BIFs* for each pixel and produce a stack of four *BIF-images* by assigning a *BIF-label* to each pixel according to the largest *BIF*; 3) encode the four *BIF-labels* of each pixel into a *BIF-column*; and 4) populate a *BIF* histogram by counting occurrences of *BIF-columns* over the whole image. As a result, all possible configurations of the seven geometric structures across the four scales comprise the visual dictionary. The *BIF-column* ranges from 1 to 7^4.

4.2.4 Textons

Texton based approaches have been widely used for texture classification where textures are modelled by the statistical distribution of a texton dictionary. A variety of textons have been developed. On the one hand, textons are generated by clustering the joint filter responses of a filter bank over a training set and the cluster centres are considered as textons. In (Varma & Zisserman, 2005), four filter sets (LM, S, MR4, and MR8) were investigated and compared (MR8 outperformed the other three). The MR8 filter bank consists of an edge and a bar filter at six orientations and three scales, a Gaussian and a Laplacian of Gaussian (LoG) filter (shown in Figure 4.4). Only eight filter responses are retained, which

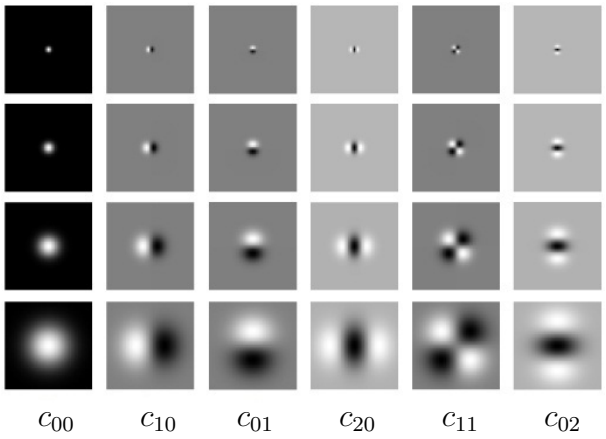


Figure 4.2: Gaussian filter bank, consisting of one zeroth-order (c_{00}), two first-order (c_{10} and c_{01}), and three second-order Gaussian derivative filters (c_{20}, c_{11} and c_{02}), each at four scales.

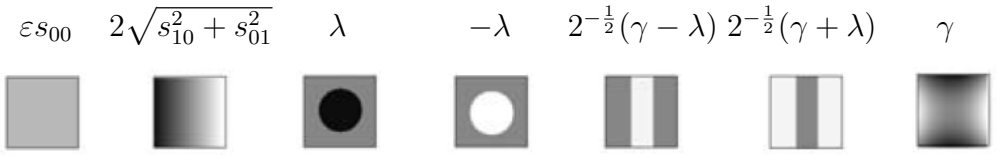


Figure 4.3: The example image patches of local geometric structures corresponding to the seven *BIFs*.

include the six maximum filter responses of the two anisotropic filters across the six orientations at each scale and the responses of the two isotropic filters. To generate texture models, a selection of training images is first chosen randomly per texture class. The joint filter responses over all the training images from one class are then aggregated and clustered using the classic k-means algorithm. The cluster centres (i.e. textons) for each class are gathered together to comprise the texton dictionary. Finally, textures are modelled by frequency histograms of textons. On the other hand, raw image patches are used to learn textons in (Varma & Zisserman, 2009). The texture modelling procedure is similar as described above. The difference is that the joint filter responses are replaced by the source image patches. Textures are modelled based on the joint distribution of original pixel values instead of filter responses of local neighbourhoods. It has been demonstrated that the image-patch based representation can provide superior performance to the filter-response based approaches.

4.3 Mammographic Tissue Appearance Modelling

Many studies have indicated that there is a strong correlation between breast tissue density/patterns and the risk of developing breast cancer. Therefore, mod-

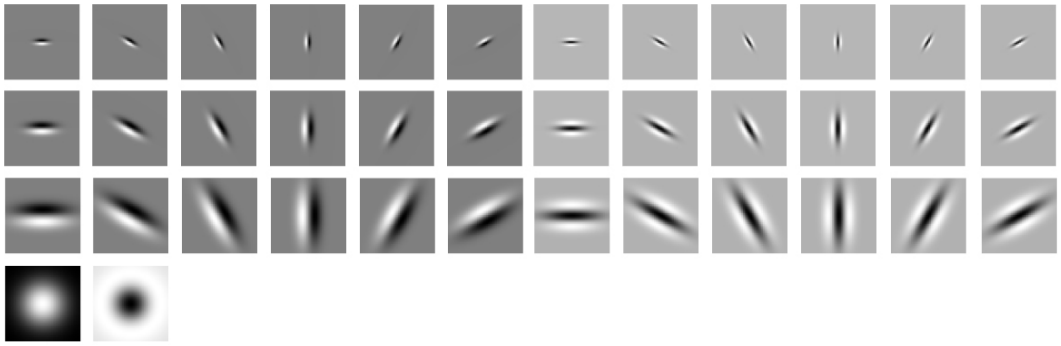


Figure 4.4: MR8 filter bank, consisting of an edge and a bar filter at six orientations and three scales, a Gaussian and a Laplacian of Gaussian filter.

elling mammographic tissue appearance is beneficial for quantitative analysis of breast density, qualitative perception of breast tissue patterns, and automated mammographic risk assessment. In this section, we present an approach to modelling breast tissue appearance in mammograms. Mammographic tissue is modelled based on statistical analysis of local tissue appearance. We investigate five strategies by employing different types of local features described in the previous section, covering local intensity, texture, and geometry information. Local features are extracted from local neighbourhoods of breast tissue pixels and are statistically analysed to build overall models of breast tissue.

The schema of our approach is shown in Figure 4.5. Firstly, as a pre-processing step, the breast region is segmented using the method described in Section 3.2. Subsequently, local features are extracted at each breast tissue pixel within the breast region (pixels in the background are excluded from feature extraction). After that, a visual dictionary is generated in the same ways as described in the previous section, containing a collection of local tissue appearance. Finally, every breast tissue pixel is labelled with the nearest "word" in the dictionary and is assigned to the corresponding histogram bin. Thus, a mammographic image is modelled as a histogram of the "words" in the dictionary.

Here, the mammographic images from the MIAS database are used for experiments to demonstrate the process of mammographic tissue appearance modelling. The original spatial resolution is $50\mu m \times 50\mu m$ per pixel. Due to memory and efficiency reasons, we down-sampled the full resolution to $800\mu m \times 800\mu m$ per pixel using a Gaussian pyramid reduction (Burt & Adelson, 1983). When modelling breast tissue appearance, to eliminate bias caused by the image edge and the breast-background boundary, we only focus on the pixels with neighbourhoods entirely located within the breast. All valid pixels can be identified automatically based on the breast mask obtained from the initial segmentation step. In addition to generating the 1-dimensional histogram models, we display the resulting models

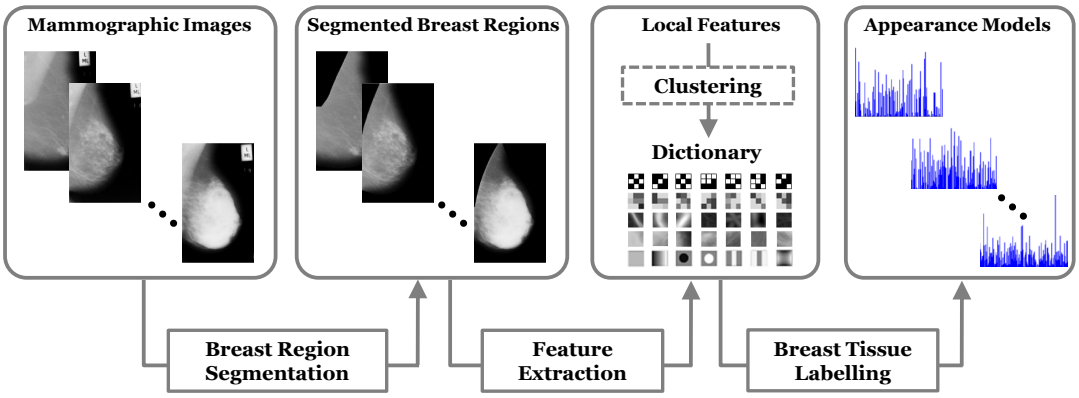


Figure 4.5: The schema of modelling breast tissue appearance based on statistical analysis of local tissue appearance.

in the form of 2-dimensional label maps, providing a visual representation of the overall distribution of local features in the breast. To generate a label map, for each breast tissue pixel, we search for the nearest (exactly matched or nearly matched) "word" to the local feature extracted at its position across the dictionary, and assign a label to it according to the response "word". The background pixels and the invalid breast tissue pixels (close to the image edge or the breast-background boundary) are set to zero. The details of the implementation of the five strategies for modelling breast tissue appearance are provided below:

- LBP

  We first use a 3×3 neighbourhood, containing 9 pixels within the local window. Thus, in Equation 4.1, the value of P is equal to 9 and the LBP number ranges from 1 to 2^9. The visual dictionary of local tissue appearance is composed of 2^9 binary patterns and the dimensionality of LBP histograms is 2^9. We also use a local window size equal to 5×5, which creates 2^{25} binary patterns. Thus, the LBP number ranges from 1 to 2^{25} and the dimensionality of the LBP histograms is 2^{25}. It should be noted that certain LBP numbers will not appear in the range given above due to the fact that the binary value of the centre pixel is always equal to 1. The redundancy can be eliminated by removing the empty histogram bins before using the LBP histograms for breast density classification. Figures 4.6 and 4.7 respectively show the resulting LBP label maps and histograms of example mammograms covering the four BIRADS categories. When using 5×5 local windows, the number of histogram bins (i.e. the feature space dimensionality) is exponentially increased, which produces a very sparse histogram. The peak occurrences in the histogram are considerably reduced.

- LGA

  As described in Section 4.2.2, there are two important parameters in the LGA based approach: the greylevel resolution and the local window size. To investigate the effect of these two parameters on the modelling results, we first use a fixed local window size of 3×3 and reduce the greylevel resolution into 8 and 16. Thus, in equation 4.2, the value of N_g is equal to 8 and 16, and the LGA number ranges from 1 to 8^9 and 16^9, respectively. The visual dictionary is composed of 8^9 and 16^9 local greylevel appearances and the dimensionality of the LGA histograms is 8^9 and 16^9, respectively. Figures 4.8 and 4.9 respectively show the resulting LGA label maps and histograms of example mammograms covering the four BIRADS categories. When using a higher greylevel resolution of 16, the LGA label maps indicate a more subtle representation of mammographic tissue. In addition, we use a fixed greylevel resolution of 16 and set the local window size equal to 3×3 and 5×5. Thus, the LGA number ranges from 1 to 16^9 and 16^{25}, respectively. Figures 4.10 and 4.11 show the resulting LGA label maps and histograms of example mammograms with respect to variation in the local window size. As shown in Figure 4.10, the LGA label maps appear almost the same for 3×3 and 5×5 local windows. It might be explained that although the data range of the colour bar is increased from $1 - 16^9$ to $1 - 16^{25}$, the relative colour values of the main components in the label maps remain almost the same. As can also be seen from Figure 4.11, the peak histogram bins appear at almost the same position with respect to the whole histogram range for 3×3 and 5×5 local windows. The LGA based approach models local tissue appearance over a range of greylevels instead of reducing local windows into binary patterns, which generates a higher-dimensionality feature space than LBP. As a result, the generated LGA histograms as shown in Figures 4.9 and 4.11 are sparser than the LBP histograms in Figure 4.7.

- BIF

  There are two parameters, ε and σ_{base}, which can be tuned when computing *BIFs*. The parameter ε determines the tolerance of a region to be considered sufficiently uniform and assigned to the first BIF category (Crosier & Griffin, 2010). The parameter σ_{base} is the finest scale of the Gaussian filter bank. To test variation of the modelling results with respect to the two parameters, we first use a fixed setting of $\varepsilon = 0$ and set σ_{base} equal to 1 and 1.2, respectively. The use of $\varepsilon = 0$ is for the purpose of mainly focusing on structure information without the flat pattern for the analysis of local tissue appearance. As described in Section 4.2.3, the seven *BIFs*

are computed at four octave-separated scales. Thus the corresponding scale range is $(1, 2, 4, 8)$ and $(1.2, 2.4, 4.8, 9.6)$, which results in 7^4 *BIF-columns*. The resulting *BIF* label maps and histograms of example mammograms are shown in Figures 4.12 and 4.13, respectively. Due to the exclusion of the flat pattern, the first *BIF* is never selected, which results in a set of *BIF-columns* never happen. As can be seen from Figure 4.13, the low dimensions of *BIF* histograms are empty (the *BIF-column* value starts from 401). In addition, we use a fixed setting of $\sigma_{base} = 1$ and set ε equal to 100 and 200, respectively. The resulting *BIF* label maps and histograms of example mammograms are shown in Figures 4.14 and 4.15, respectively. When using a large ε value of 200, dark blue filters into the breast, which indicates low-contrast structures are smoothed and assigned to the flat category. The distribution of *BIF-columns* is extended to the whole histogram (some peak histogram bins are present at the low dimensions). As shown in Figures 4.12 and 4.13, the resulting *BIF* label maps do not seem to correlate with breast tissue density. However, it should be noted that the *BIF* label map aims to describe the aspect of geometric structure of local tissue appearance, where different colours represent different configurations of the seven *BIFs* across the four scales, and therefore there is no direct correlation between the colour of a pixel in the *BIF* label map and its original greylevel value. Different arrangement of the seven *BIFs* can change the colour value of a *BIF-column*.

- Texton I and Texton II

  We use both types of textons as described in Section 4.2.4 for modelling mammographic tissue appearance. We refer to the two modelling strategies based on the MR8 and image-patch textons as Texton I and Texton II, respectively. We randomly selected 40 images from the MIAS database as the training set to learn the textons of breast tissue appearance in mammograms. When selecting the training images, BIRADS categories of mammograms are not taken into account (we do not make an implicit assumption that mammograms in each BIRADS class have the same tissue appearance). For Texton I, we set $(\sigma_x, \sigma_y) = \{(1, 3), (2, 6), (4, 12)\}$ for the oriented filters and set $\sigma = 10$ for the Gaussian and LoG filters in the MR8 filter bank. After filtering all training images with the MR8 filter bank, a set of 8 dimensional vectors of filter responses are produced. For Texton II, 3×3 image patches are extracted around each valid pixel (i.e. its local neighbourhood is entirely located within the breast region) from all training images, and then these extracted image patches are rearranged in row order to form a set of 9 dimensional vectors of raw image pixels. Subsequently, the classic

k-means algorithm is employed to cluster the resulting 8 (9) dimensional vectors and generate the textons. For each individual texton type, 40, 80 and 160 textons are generated and sorted according to their magnitude values in ascending order. Thus, the dictionary is composed of 40, 80 and 160 textons and the dimensionality of texton histograms is 40, 80 and 160, respectively. Finally, the label map is obtained by labelling each breast tissue pixel with the nearest texton. Note that the pre-processing step in (Varma & Zisserman, 2005, 2009) is not applied here, and the filter responses in Texton I are not normalised by Weber's law as in (Varma & Zisserman, 2005, 2009), in order to retain the original intensity correlation between mammograms. Figures 4.16 and 4.17 show the resulting label maps and texton histograms for Texton I when generating 40, 80 and 160 textons. Figures 4.18 and 4.19 show the resulting label maps and texton histograms for Texton II when generating 40, 80 and 160 textons. As shown in Figures 4.16 and 4.18, the label maps for each example mammogram have an approximately similar appearance when using different numbers of textons. As the number of textons increases, more local details are retained due to the increase in resolution. The texton histograms generated using a larger number of textons are more evenly distributed (Figures 4.17 and 4.19).

As indicated from the resulting label maps of the five modelling strategies, LBP is sensitive to noise and small textures, and therefore no large homogeneous regions are obtained within the breast region. The label maps for LGA and Texton II indicate a high correlation with respect to breast tissue density. The area of dense tissue regions (labelled with red colour) in mammograms increases from BIRADS I to BIRADS IV. Texton I has strong responses to the boundary between dense and fatty tissue regions, and relatively homogeneous regions are obtained within the dense/fatty tissue. For BIF, cell-like regions are obtained within the breast, representing the mixture of geometric structures across multiple scales. For the following experiments and further evaluation provided in subsequent chapters, the parameters in the five algorithms are set as below: 3×3 neighbours for LBP, LGA and Texton II, $N_g = 16$, $\varepsilon = 0$, $\sigma_{base} = 1$, $(\sigma_x, \sigma_y) = \{(1,3),(2,6),(4,12)\}$, $\sigma = 10$, and 160 textons for Texton I and Texton II.

As described above, the occurrence histograms of the five types of local features are populated by spanning the full dictionary. This could result in sparse histograms as some image appearances never happen in the real observation of breast tissue. To investigate all possible occurrences of the five types of local features in the MIAS database, five occurrence histograms are generated across the full database, each corresponding to one type of local features (Figure 4.20 (left column)). It is

shown that many histogram bins are empty or nearly empty especially in higher-dimensional histograms. This indicates that many "words" in the dictionary are rarely referenced. Therefore, we can use only the informative part instead of the full histograms for the classification. To select the dominant "words" from the dictionary for each type of local features, we sort the histogram bins of each occurrence histogram in descending order and as such the front histogram bins indicate the occurrences of the most frequently used "words". We then calculate an accumulative histogram for each sorted occurrence histogram and choose the "words" of which corresponding occurrences occupy a high percentage (varying from 60% to 99% for different approaches or parameter values in this work) of the total occurrence as the dominant "words". Thus, the "words" which are never or rarely referenced are removed and the retained "words" comprise a more common dictionary. It is shown by experiments that using the frequently referenced "words" can effectively retain the fineness of the histogram representation and remove noise (i.e. never/rarely referenced "words") simultaneously. Relatively compact histograms with lower dimensions can be obtained by referring to the more common dictionary. The regenerated five occurrence histograms after compressing the visual dictionary are shown in Figure 4.20 (right column) where the dictionary is composed of the "words" which make up 99% occurrences. It is shown that the dimensionality of the occurrence histograms is significantly reduced after compressing the visual dictionary, especially for LBP, LGA, and BIF. As we use a high percentage of 99% here, the removed histogram bins are mainly those empty or almost empty ones and as such there is no significant change in the layout of the histograms. The resulting common dictionary for each type of local features are used to generate the results provided in subsequent chapters.

The modelling results of breast tissue appearance can be used for breast density classification. The resulting models in the form of occurrence histograms are used as feature vectors for classification. Further evaluation for mammographic risk assessment based on the full MIAS database will be provided in Chapter 6.

4.4 Summary

In this chapter, a brief review of texture analysis has been presented, which concentrates on various methods for texture feature extraction, categorised as statistical, structural, signal processing, and model based methods. Different approaches to representing texture images based on local features have been investigated and compared. On the basis of that, we presented an overall framework for modelling breast tissue in mammograms based on statistical analysis of local tissue appearance. We generated a visual dictionary of generic breast tissue appearance by

aggregating local features extracted from mammographic images. We exploited five types of local features and proposed five strategies for modelling mammographic tissue appearance. In addition, we investigated the modelling results with respect to variation in parameters. To our knowledge, this work is the first attempt to make use of the five texture representations in the field of mammographic image analysis. The obtained label maps have been used for qualitative evaluation, which indicate various representations of breast tissue patterns in the aspects of intensity, texture and geometry. For quantitative evaluation, the resulting histogram models will be applied to breast density classification for the purpose of mammographic risk assessment in Chapter 6.

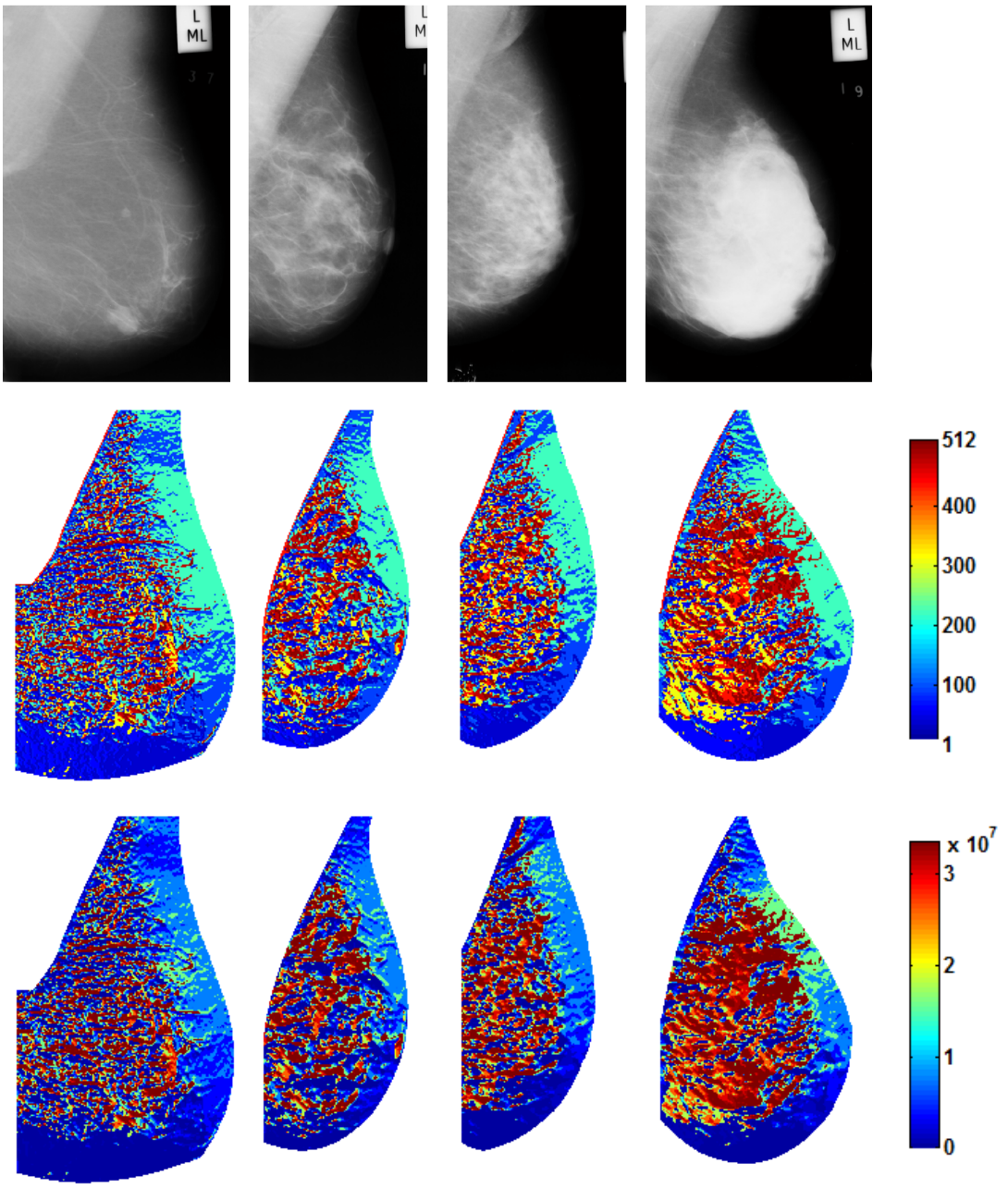
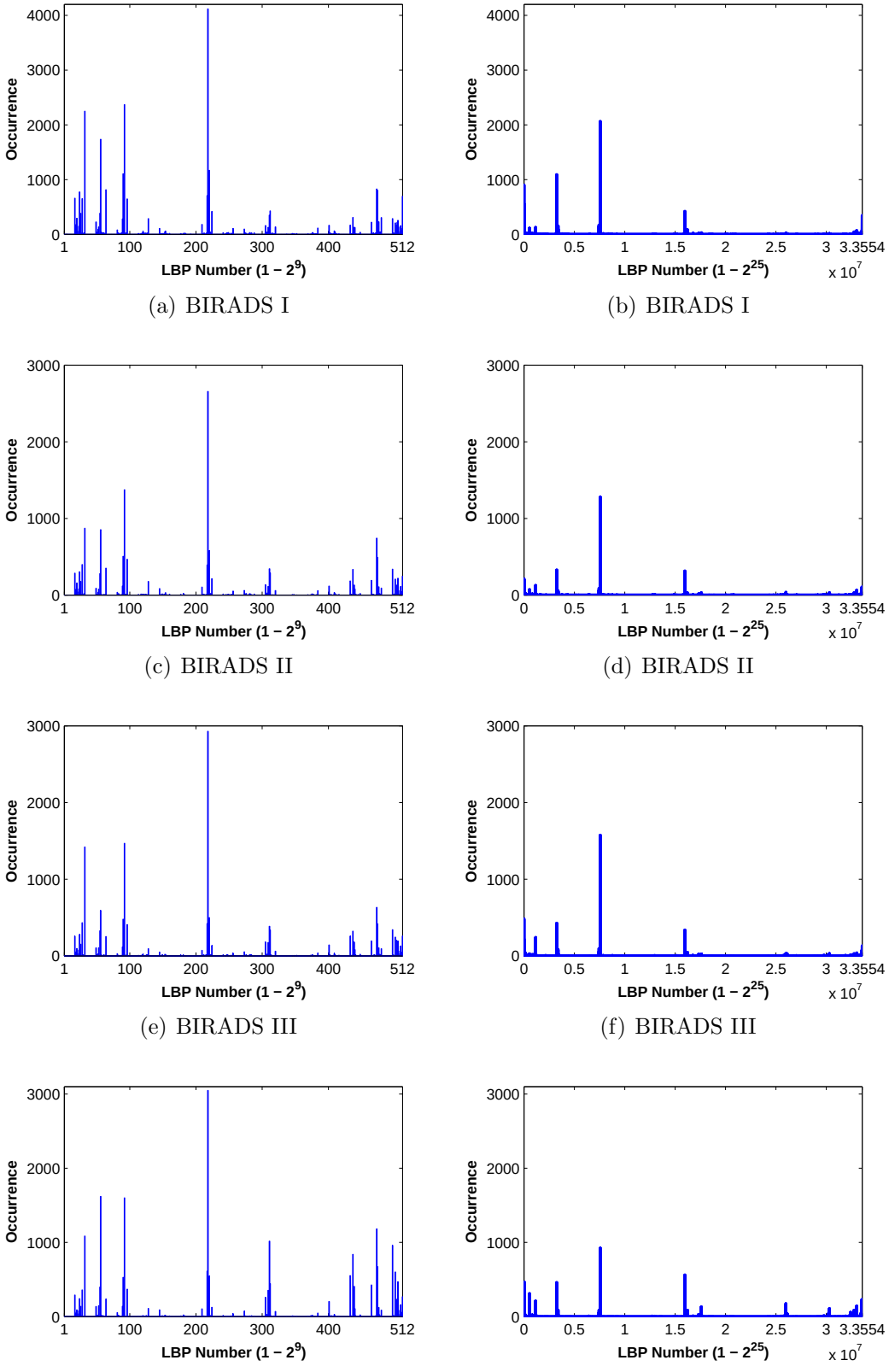


Figure 4.6: Example mammograms (top row) and the resulting *LBP* label maps when using 3×3 (middle row) and 5×5 (bottom row) local windows. From left to right, the mammograms are sorted from BIRADS I to BIRADS IV. Different colours show different local binary patterns. A color bar is given to show the colour data range.



(a) BIRADS I (b) BIRADS I

(c) BIRADS II (d) BIRADS II

(e) BIRADS III (f) BIRADS III

(g) BIRADS IV (h) BIRADS IV

Figure 4.7: The *LBP* histograms of the four example mammograms when using 3×3 (left column) and 5×5 (right column) local windows.

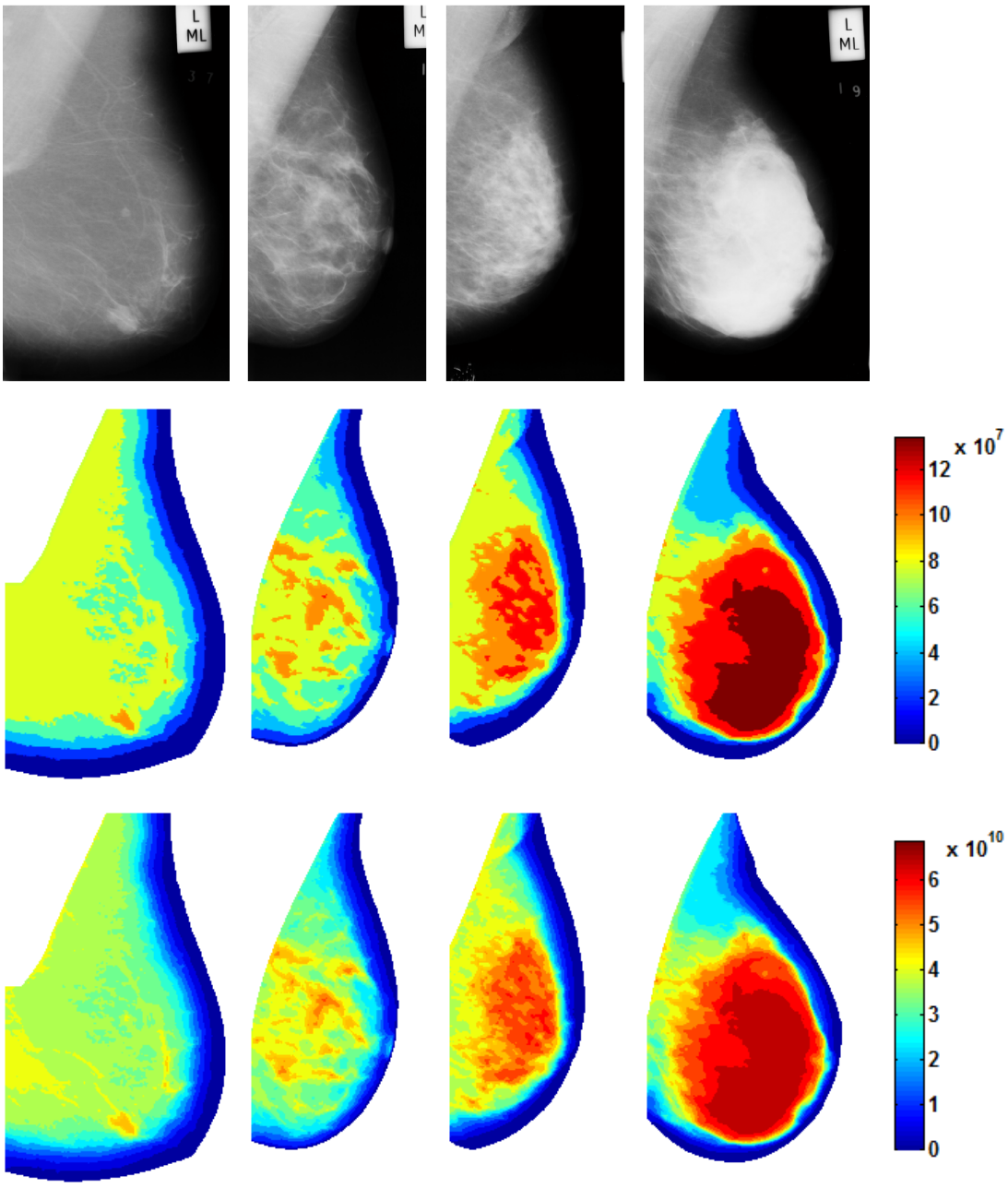
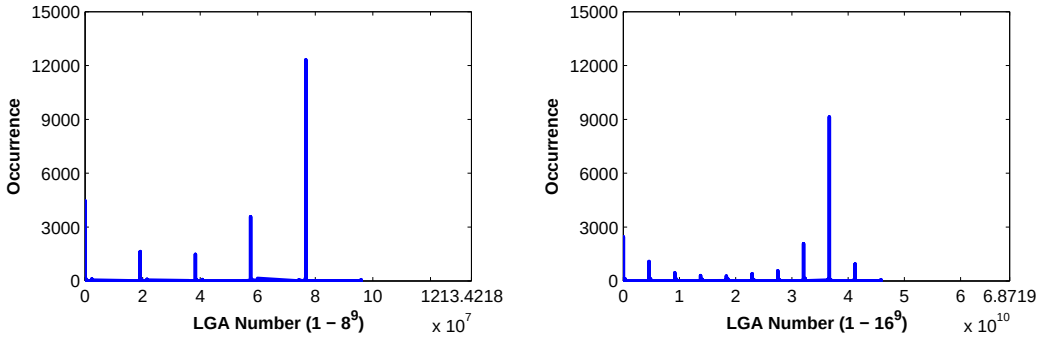
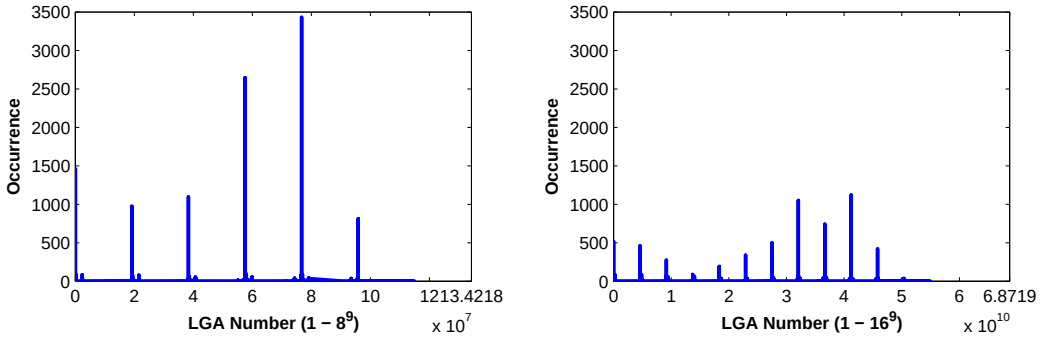


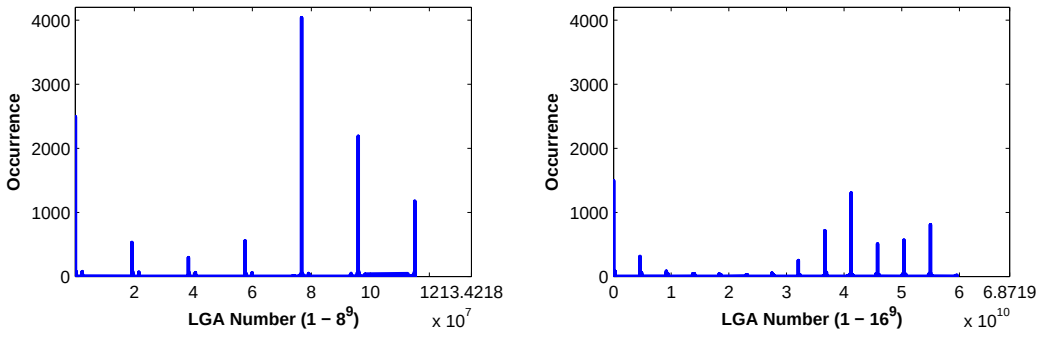
Figure 4.8: Example mammograms (top row) and the resulting *LGA* label maps when using a greylevel resolution of 8 (middle row) and 16 (bottom row). The local window size is equal to 3×3. From left to right, the mammograms are sorted from BIRADS I to BIRADS IV. Different colours show different local greylevel appearances. A color bar is given to show the colour data range.



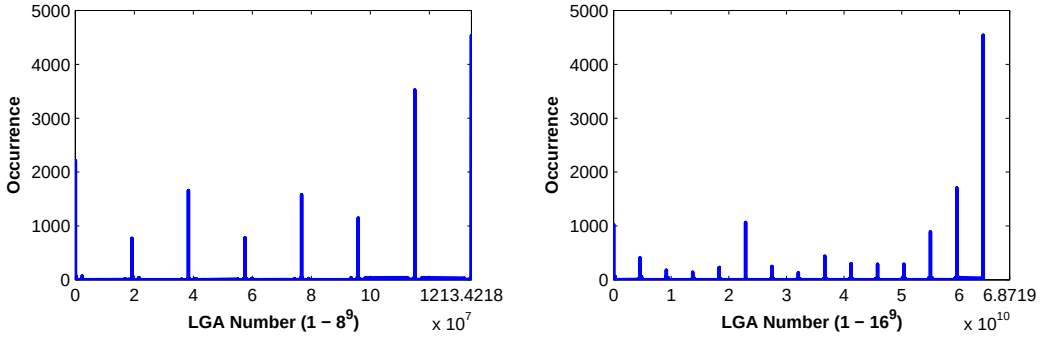
(a) BIRADS I (b) BIRADS I



(c) BIRADS II (d) BIRADS II



(e) BIRADS III (f) BIRADS III



(g) BIRADS IV (h) BIRADS IV

Figure 4.9: The *LGA* histograms of the four example mammograms when using a greylevel resolution of 8 (left column) and 16 (right column). The local window size is equal to 3×3.

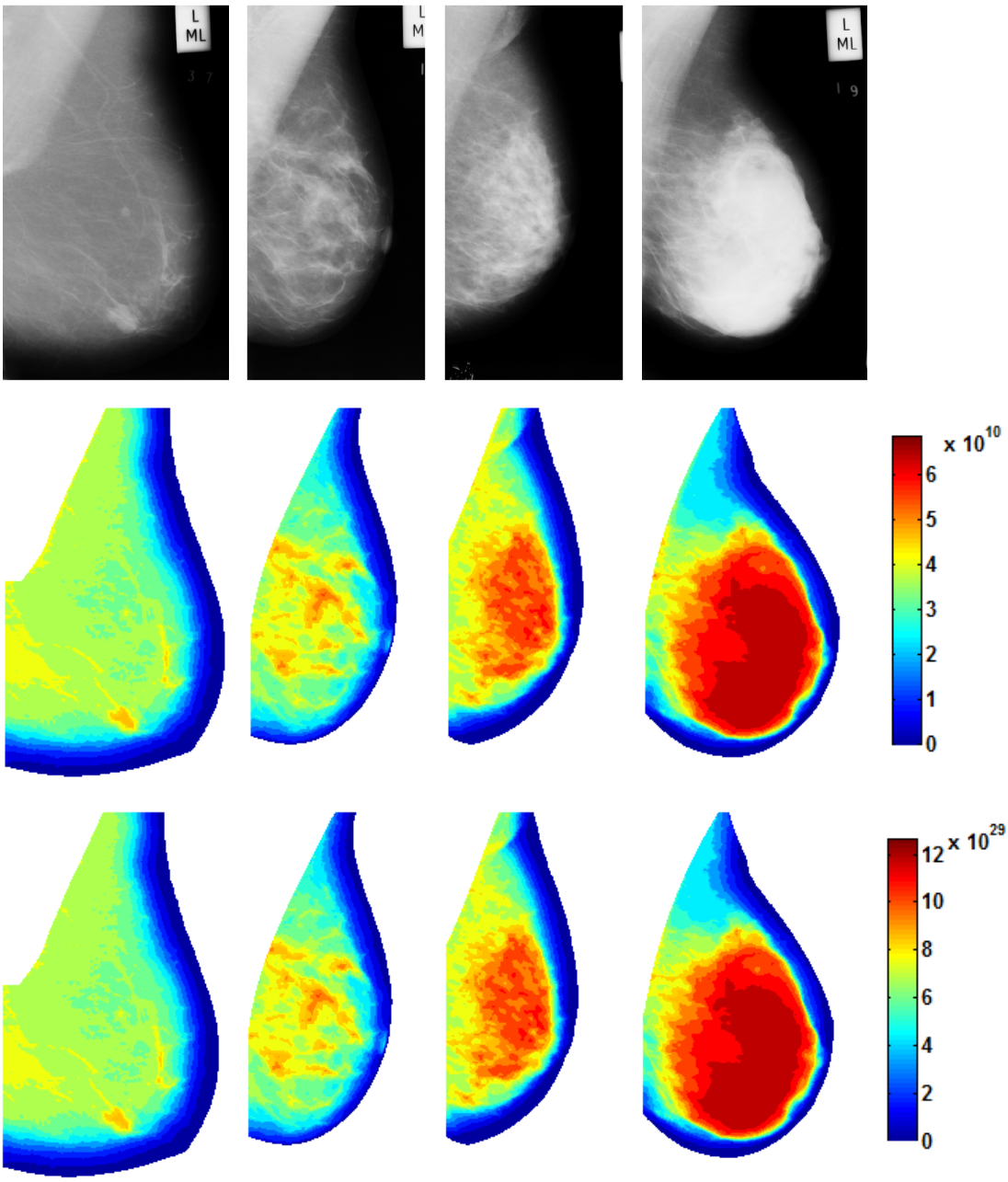
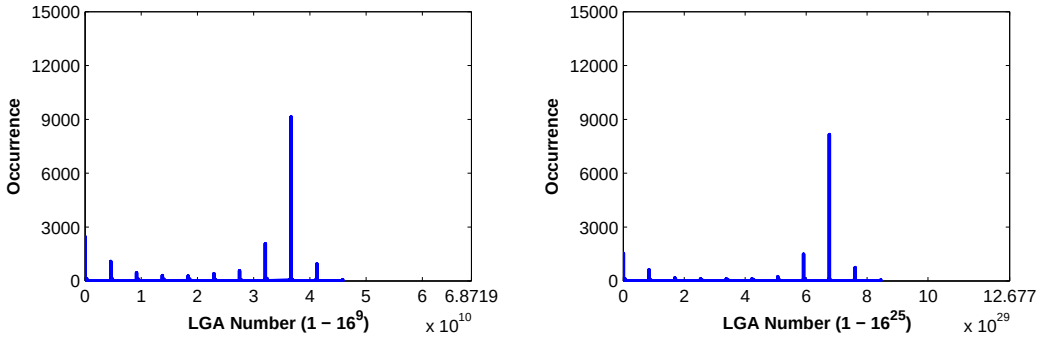
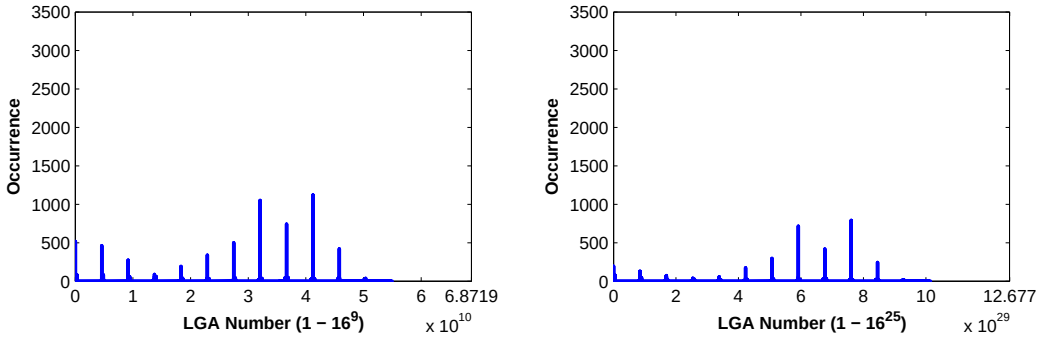


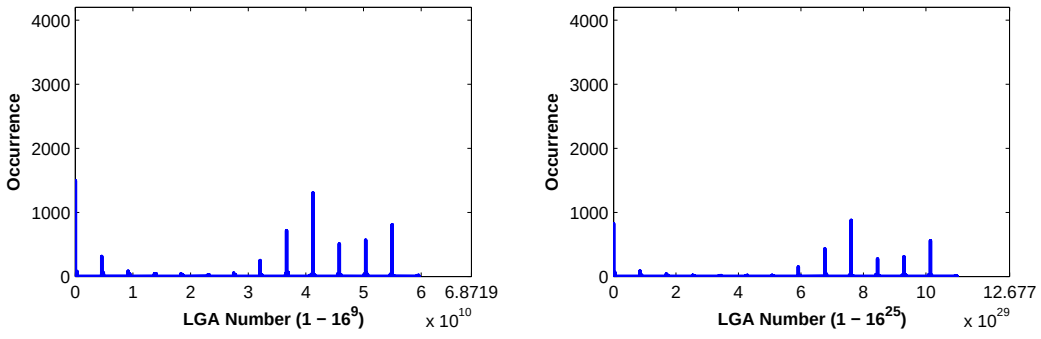
Figure 4.10: Example mammograms (top row) and the resulting LGA label maps when using 3×3 (middle row) and 5×5 (bottom row) local windows. The greylevel resolution is equal 16. From left to right, the mammograms are sorted from BIRADS I to BIRADS IV. Different colours show different local greylevel appearances. A color bar is given to show the colour data range.



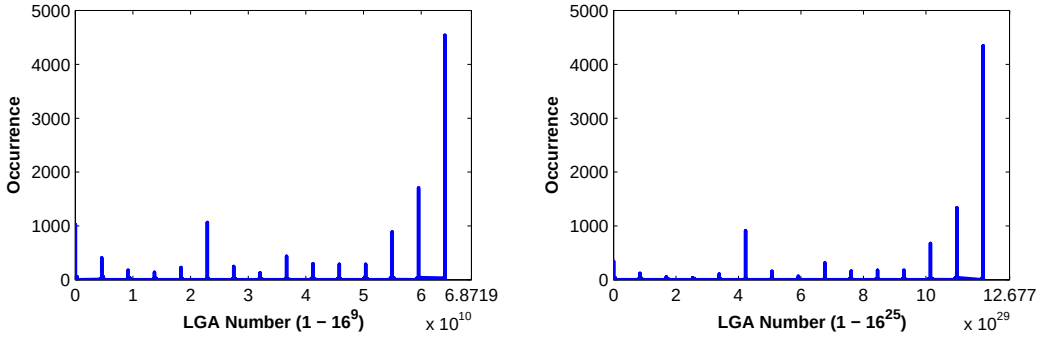
(a) BIRADS I (b) BIRADS I



(c) BIRADS II (d) BIRADS II



(e) BIRADS III (f) BIRADS III



(g) BIRADS IV (h) BIRADS IV

Figure 4.11: The *LGA* histograms of the four example mammograms when using 3×3 (left column) and 5×5 (right column) local windows. The greylevel resolution is equal 16.

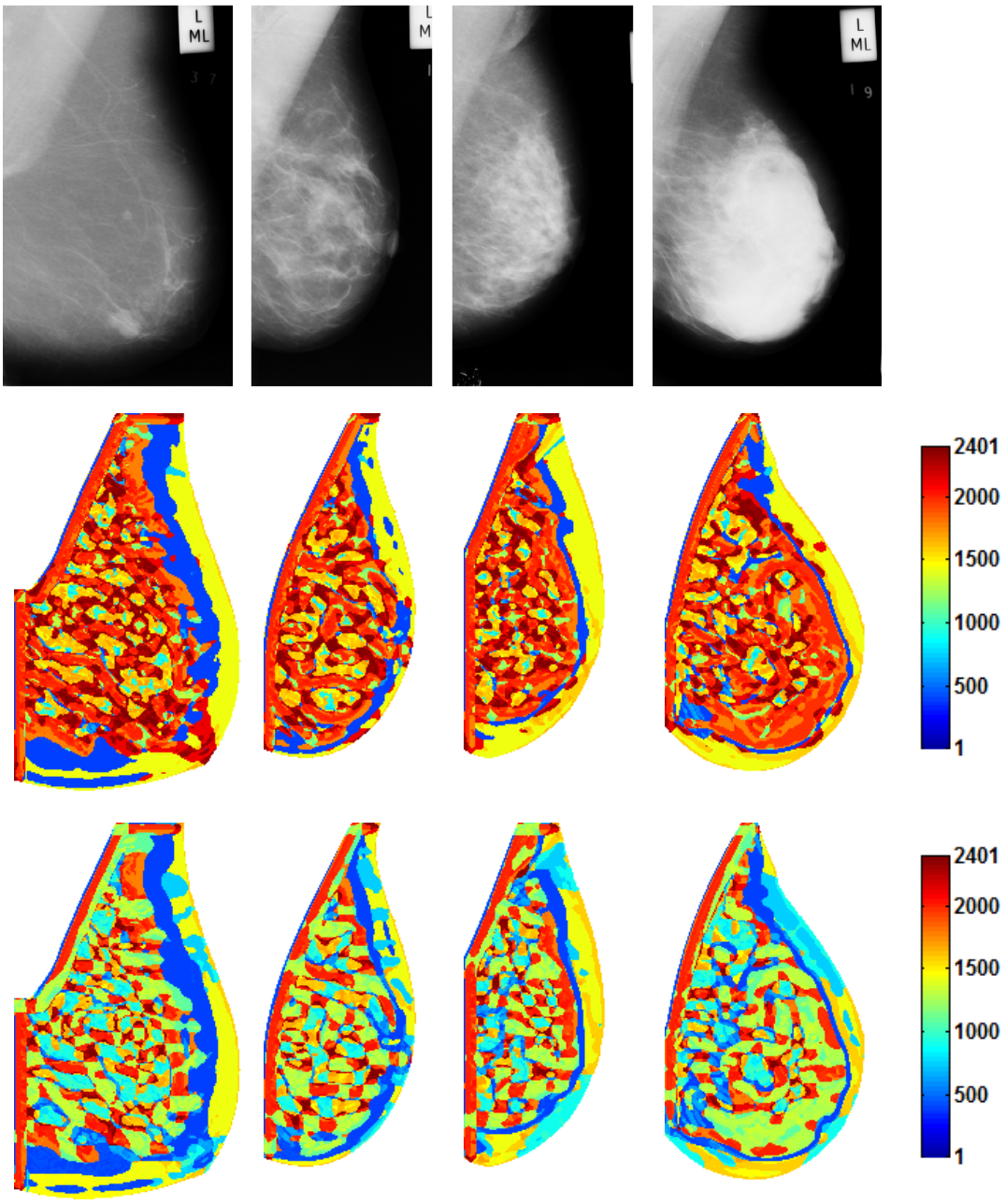
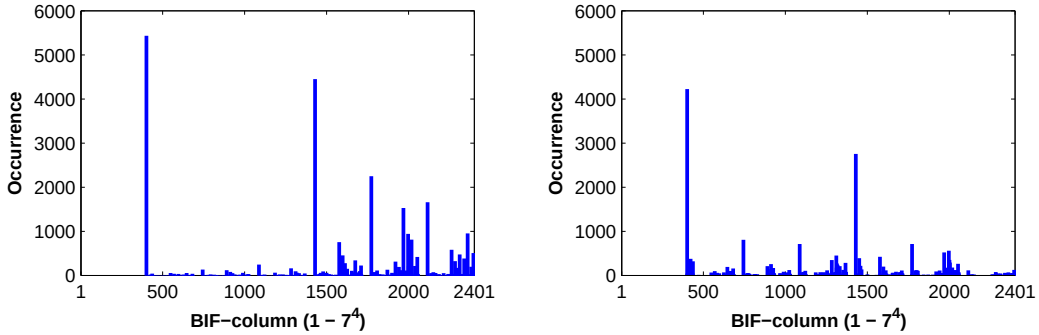
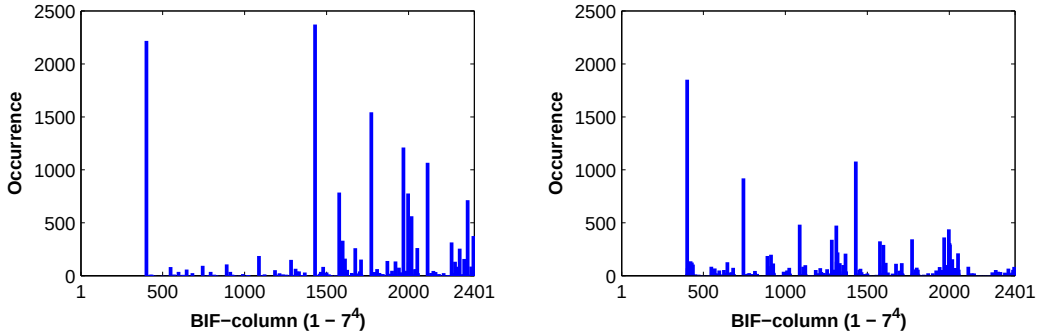


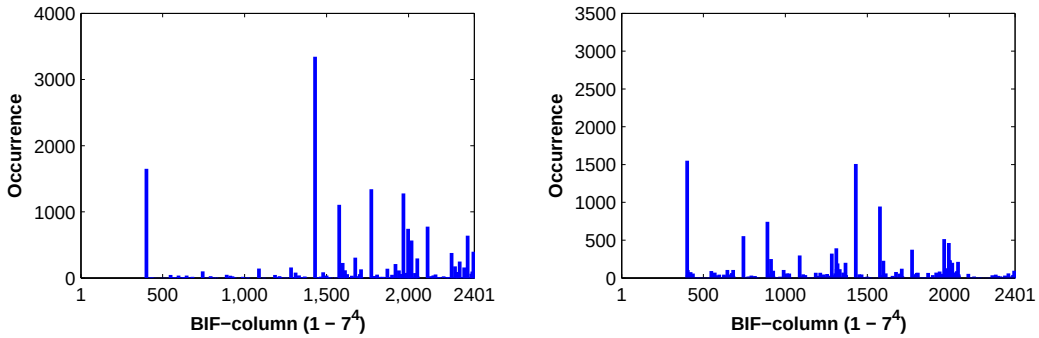
Figure 4.12: Example mammograms (top row) and the resulting *BIF* label maps when using a fixed value of $\varepsilon = 0$ and setting $\sigma_{base} = 1$ (middle row) and $\sigma_{base} = 1.2$ (bottom row). From left to right, the mammograms are sorted from BIRADS I to BIRADS IV. Different colours represent different configurations of a stack of *BIFs* over four scales. A color bar is given to show the colour data range.



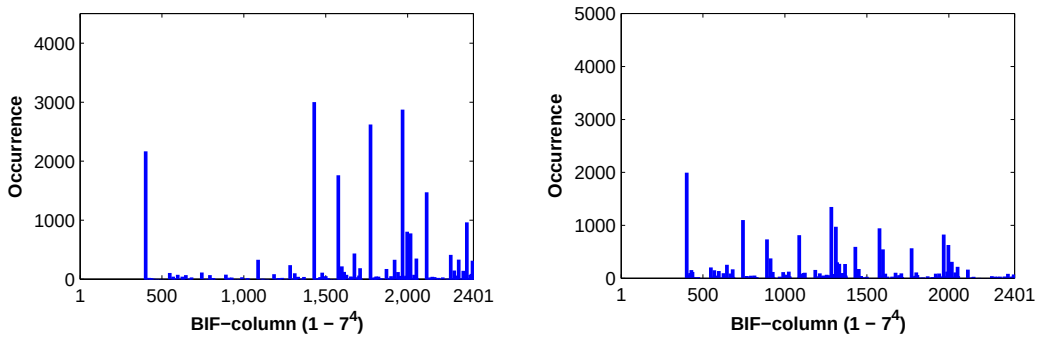
(a) BIRADS I (b) BIRADS I



(c) BIRADS II (d) BIRADS II



(e) BIRADS III (f) BIRADS III



(g) BIRADS IV (h) BIRADS IV

Figure 4.13: The *BIF* histograms of the four example mammograms when using a fixed value of $\varepsilon = 0$ and setting $\sigma_{base} = 1$ (left column) and $\sigma_{base} = 1.2$ (right column).

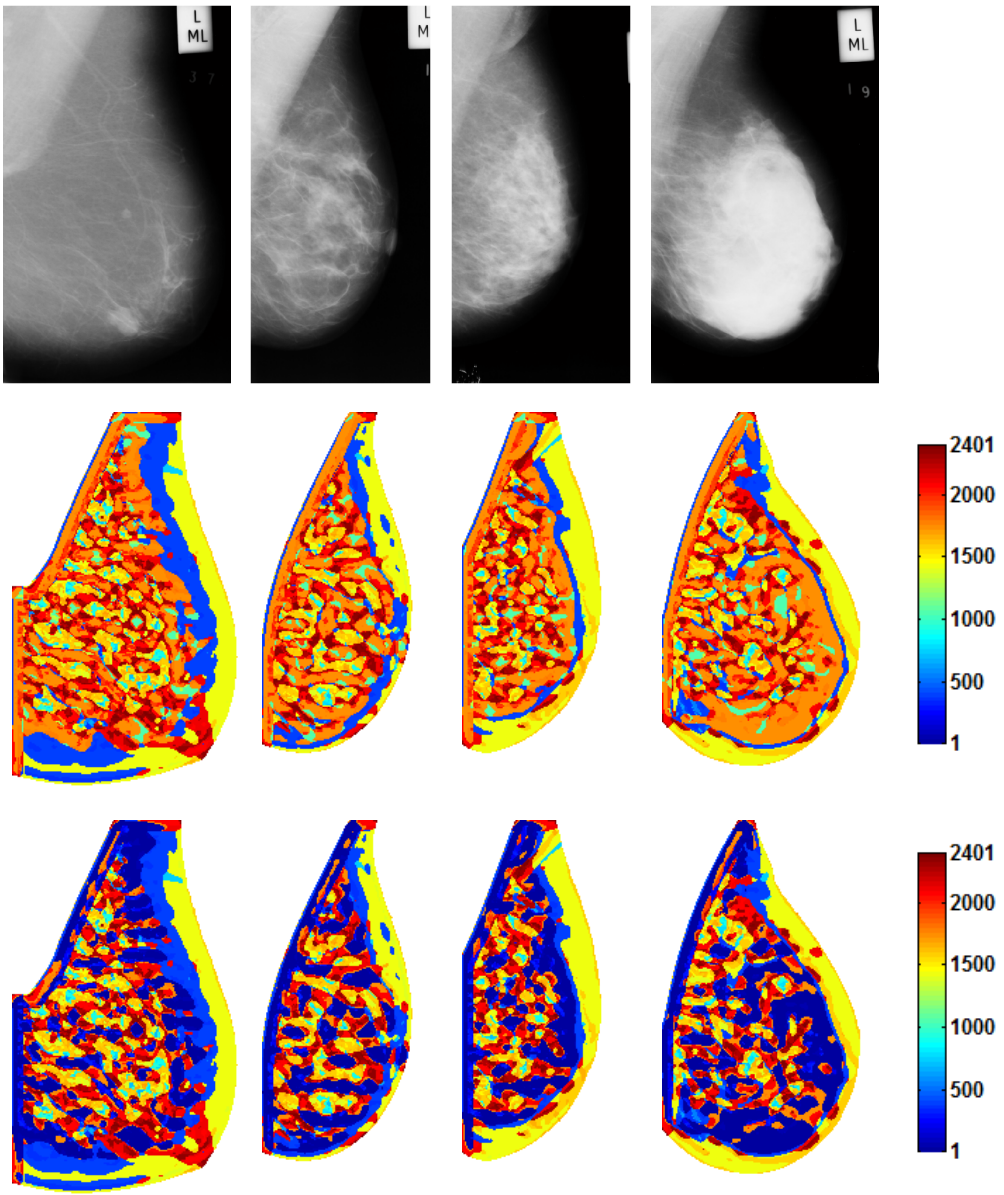
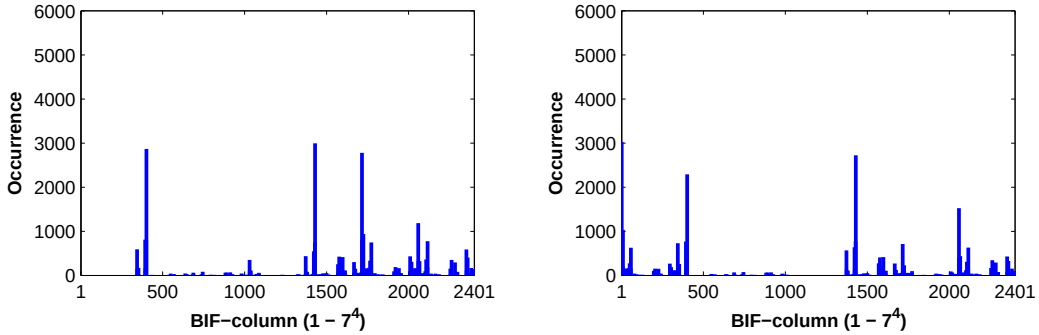
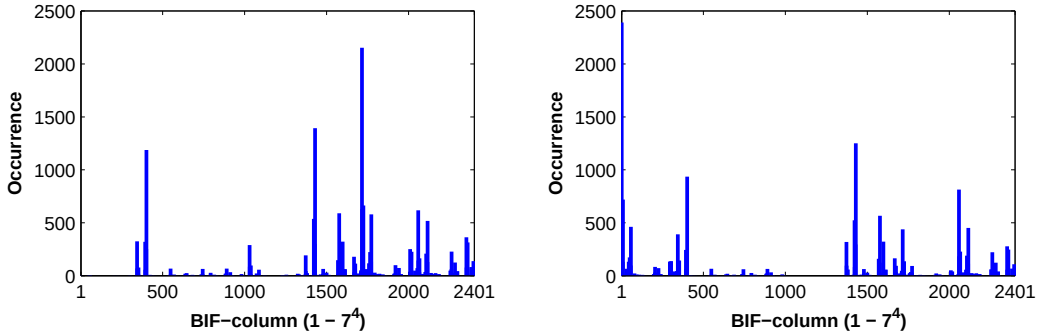


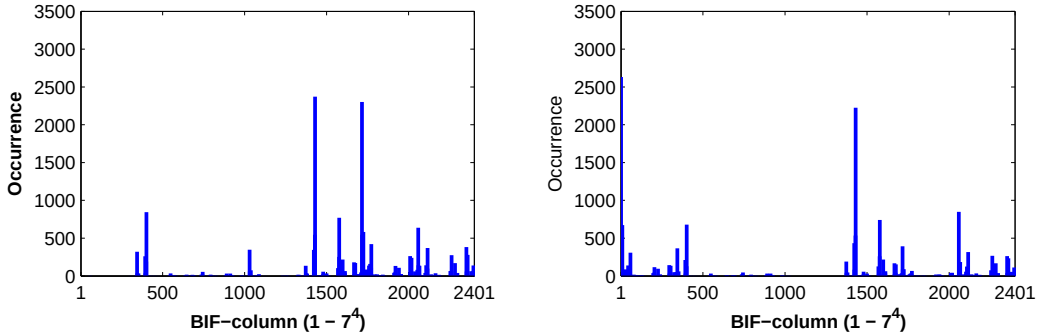
Figure 4.14: Example mammograms (top row) and the resulting *BIF* label maps when using a fixed valued of $\sigma_{base} = 1$ and setting $\varepsilon = 100$ (middle row) and $\varepsilon = 200$ (bottom row). From left to right, the mammograms are sorted from BIRADS I to BIRADS IV. Different colours represent different configurations of a stack of *BIFs* over four scales. A color bar is given to show the colour data range.



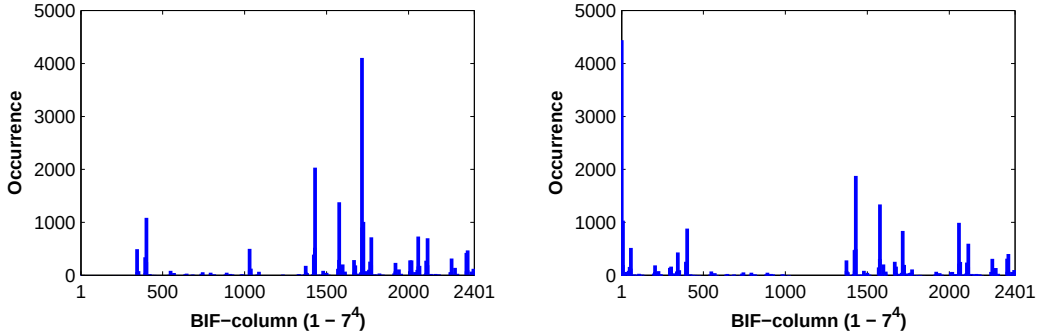
(a) BIRADS I (b) BIRADS I



(c) BIRADS II (d) BIRADS II



(e) BIRADS III (f) BIRADS III



(g) BIRADS IV (h) BIRADS IV

Figure 4.15: The *BIF* histograms of the four example mammograms when using a fixed value of $\sigma_{base} = 1$ and setting $\varepsilon = 100$ (left column) and $\varepsilon = 200$ (right column).

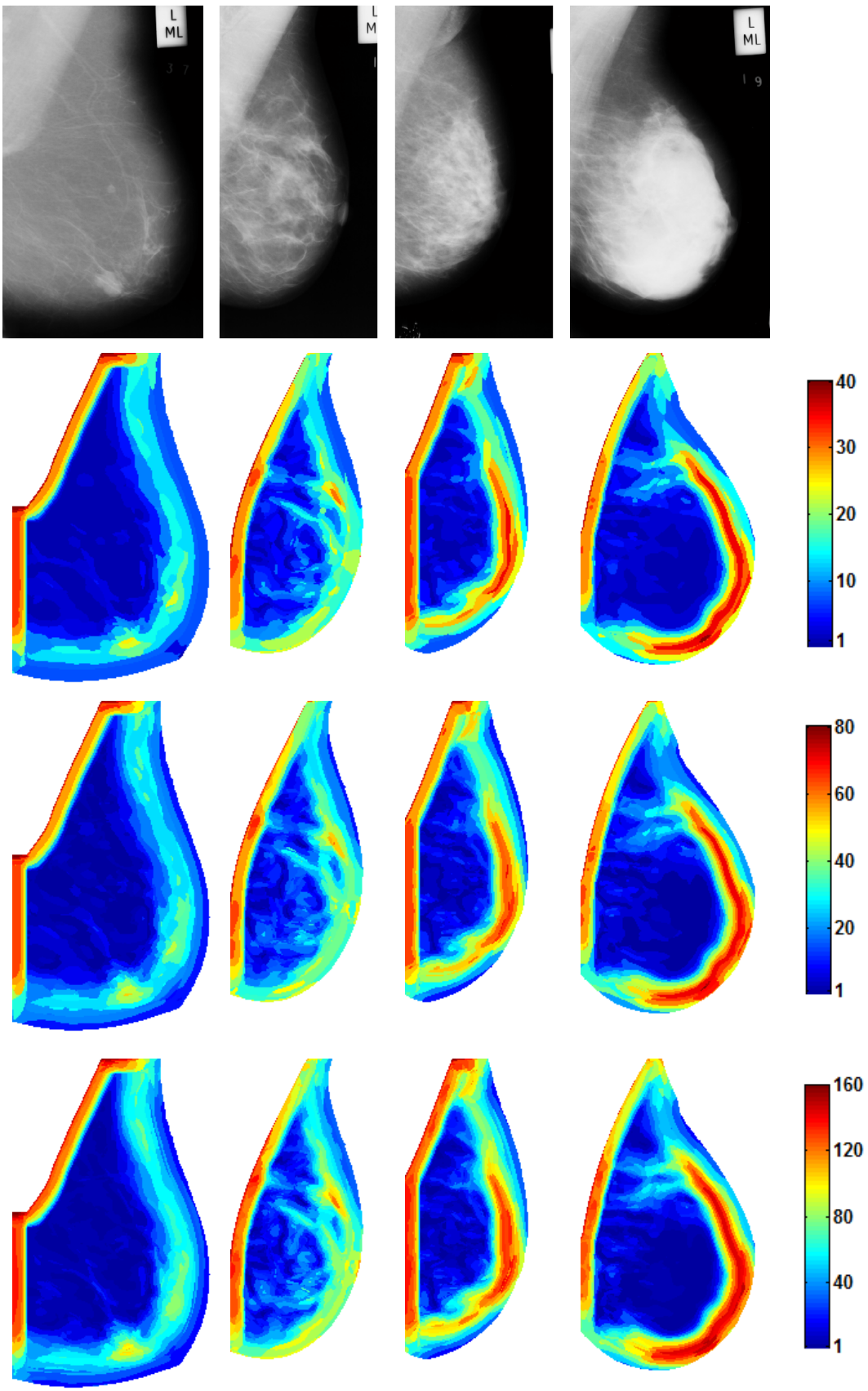
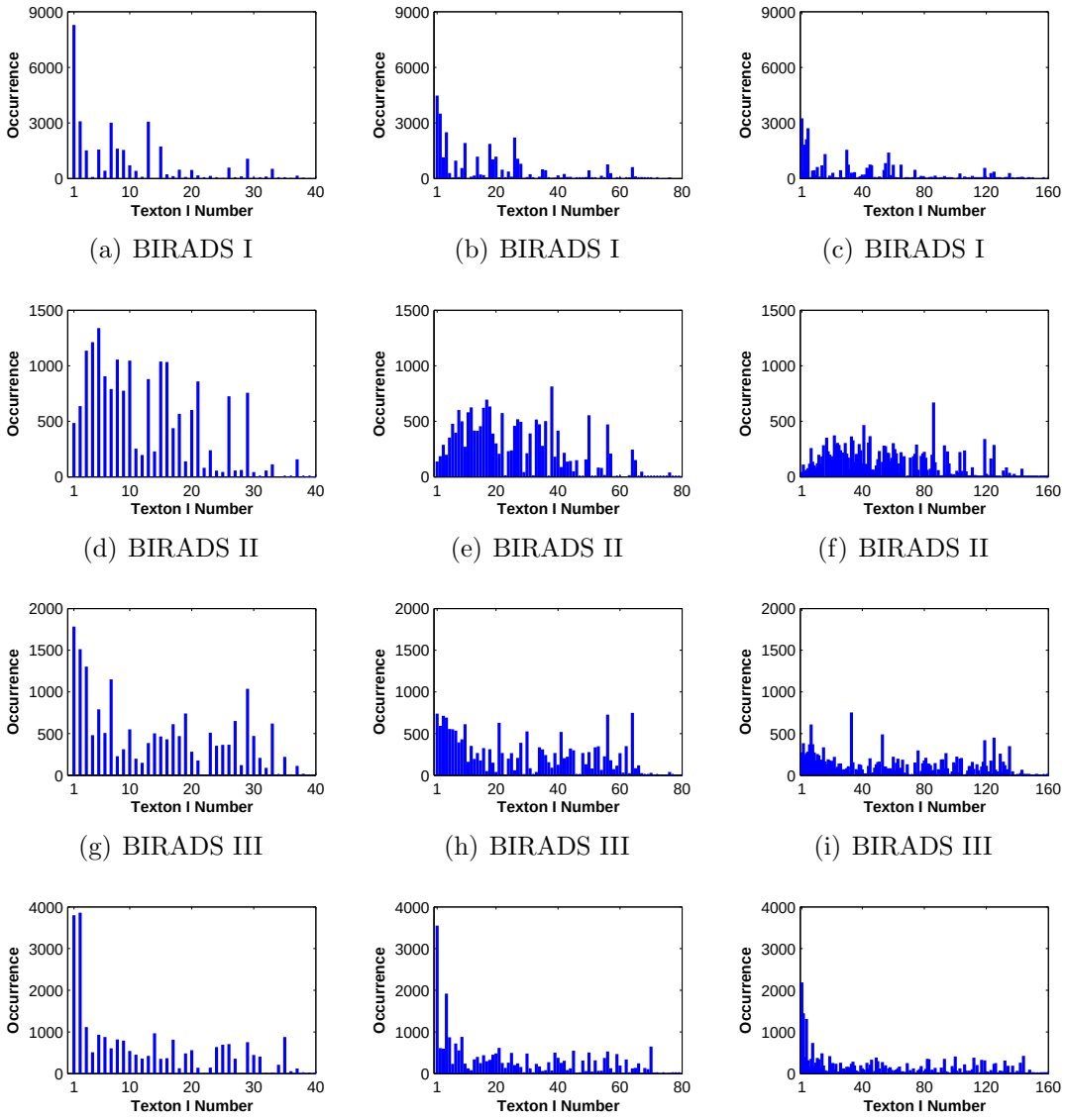


Figure 4.16: Example mammograms (top row) and the resulting Texton I label maps when generating 40 (second row), 80 (third row), and 160 (bottom row) textons. From left to right, the mammograms are sorted from BIRADS I to BIRADS IV. Different colours represent different textons.



(a) BIRADS I (b) BIRADS I (c) BIRADS I

(d) BIRADS II (e) BIRADS II (f) BIRADS II

(g) BIRADS III (h) BIRADS III (i) BIRADS III

(j) BIRADS IV (k) BIRADS IV (l) BIRADS IV

Figure 4.17: The texton histograms of the four example mammograms for Texton I when using 40 (left column), 80 (middle column), and 160 (right column) textons.

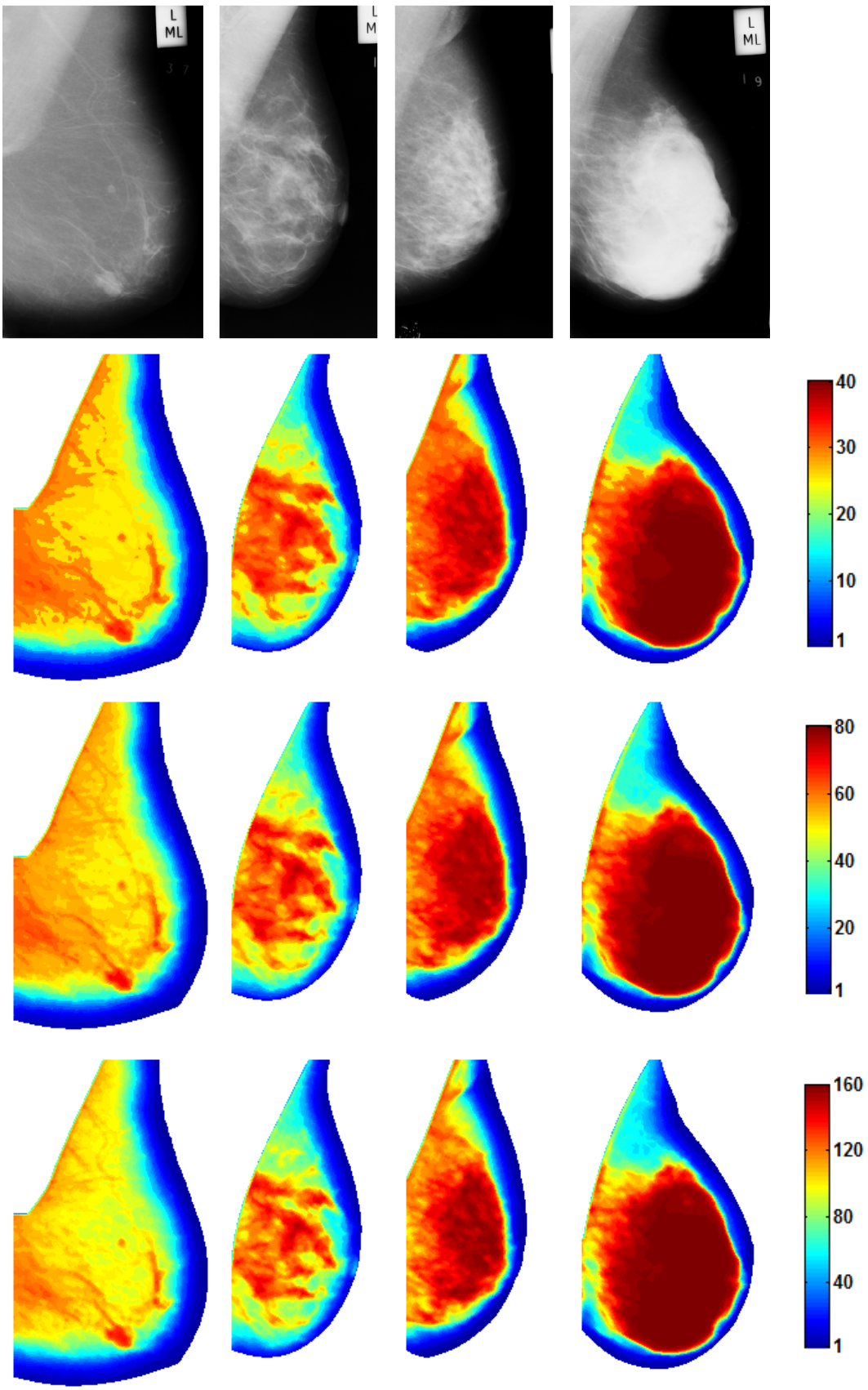
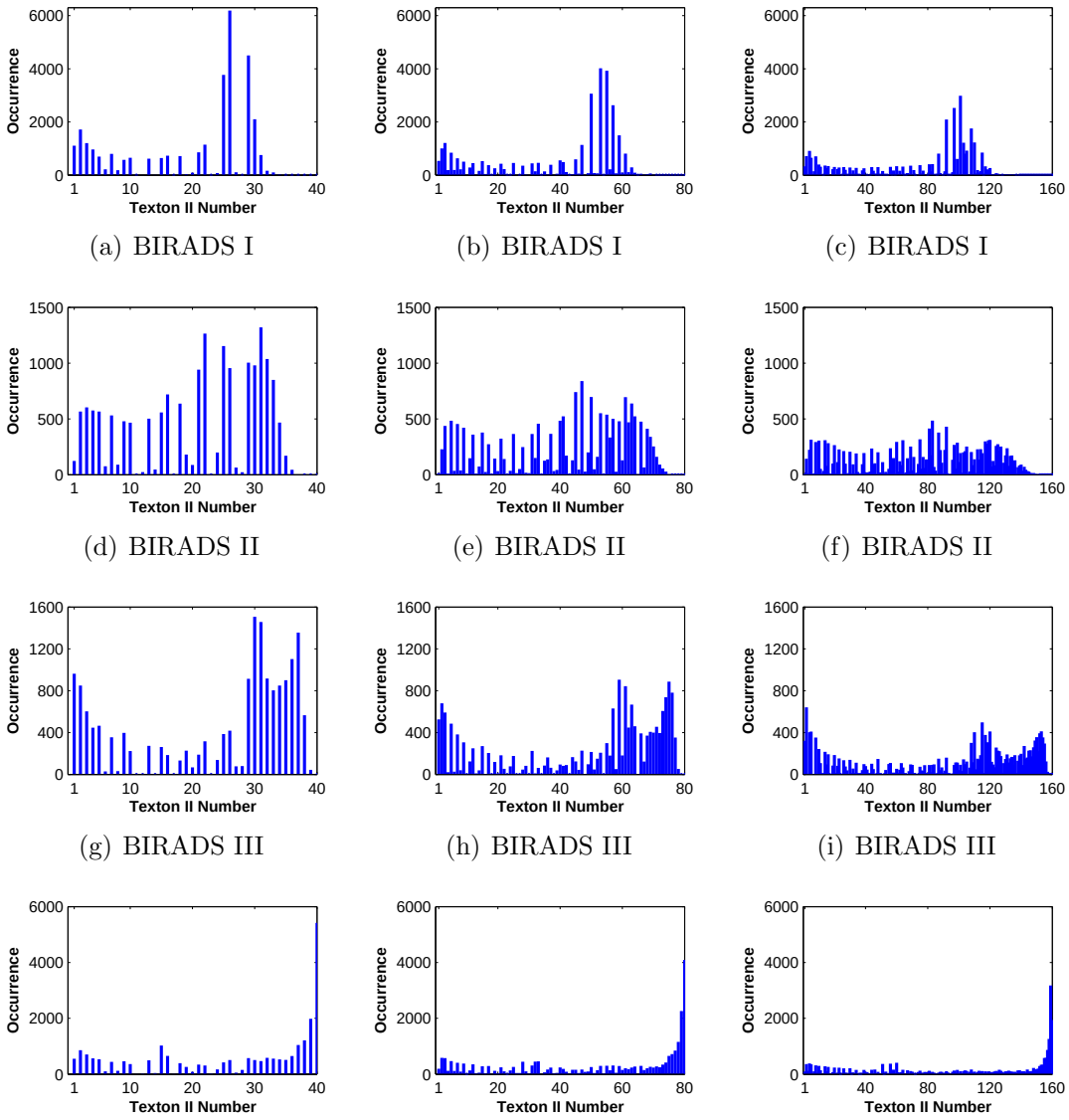


Figure 4.18: Example mammograms (top row) and the resulting Texton II label maps when generating 40 (second row), 80 (third row), and 160 (bottom row) textons. From left to right, the mammograms are sorted from BIRADS I to BIRADS IV. Different colours represent different textons.



(a) BIRADS I (b) BIRADS I (c) BIRADS I

(d) BIRADS II (e) BIRADS II (f) BIRADS II

(g) BIRADS III (h) BIRADS III (i) BIRADS III

(j) BIRADS IV (k) BIRADS IV (l) BIRADS IV

Figure 4.19: The texton histograms of the four example mammograms for Texton II when using 40 (left column), 80 (middle column), and 160 (right column) textons.

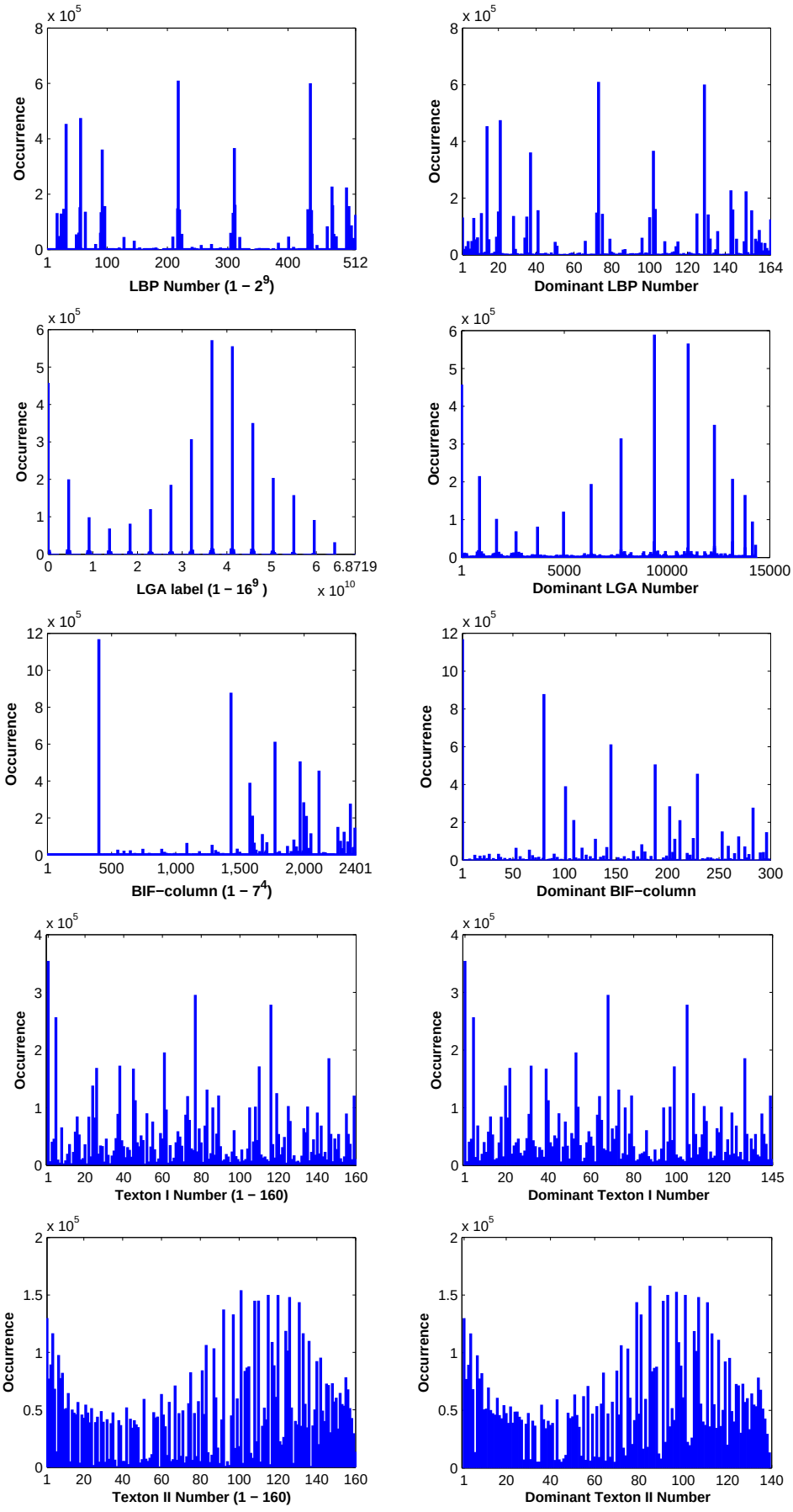


Figure 4.20: The occurrence histograms generated across the MIAS database using the whole dictionary (left column) and the compressed dictionary (right column).

Chapter 5

Detection of Blob-like Image Structures in Scale Space

Blob-like structures are one of the most distinctive structures in an image. Unlike edges and corners, blobs can provide complementary information about regions. The detection of blob-like image structures can be linked to many well-known techniques for detecting interest points/regions, which have been widely applied to various tasks in computer vision. The location of a blob-like structure can be identified by searching for scale-space extrema, and its characteristic size is indicated by the scale level at which an extremum is detected over scales (Lindeberg, 1998; Mikolajczyk & Schmid, 2004). In this chapter, firstly, an introduction of scale-space theory is provided, which describes a general idea of multiscale image representation. Subsequently, a set of blob detectors (which can also be referred to as interest point detectors) are briefly reviewed, which can determine the spatial location and the scale of blobs simultaneously. After that, a blob based representation of mammographic parenchymal patterns is proposed. We focus on approximately blob-like dense tissue patterns (e.g. nodular and homogeneous tissue) to model parenchymal patterns of a breast as a set of multiscale blobs. We detect bright blobs at multiple scales, corresponding to blob-like dense tissue regions of various sizes. Qualitative relations among blobs are incorporated to remove the overlapping and form a concise blob representation. The distribution of blobs over multiple scales can be deemed to be a general description of breast density. A measure of breast density is defined based on the proposed blob representation, which will be used for mammographic risk assessment in Chapter 6.

5.1 Scale-Space Theory

Scale-space theory provides a framework for multiscale image representation, which has been widely applied in computer vision. Its core is to deal with the multiscale nature of objects in the real world. The perception of objects may be different, depending on the scale of observation. When interpreting an image, there is generally no prior knowledge of the size of objects in the image. Therefore, it is essential to analyse the image representation at multiple scales simultaneously. A scale-space representation of an image is achieved by embedding the image into a one-parameter family of smoothed images, where fine-scale structures are progressively suppressed (Koenderink, 1984; Lindeberg, 1994, 2008). The parameter in this family is referred to as the scale parameter, which determines the fineness of structures retained in the image representation at this scale level.

When generating such a multiscale representation of an image, an important prerequisite is that the representations at coarse scales should constitute simplifications of corresponding structures at finer scales. New structures should not be created in the transformation from fine to coarse scales (Lindeberg, 1994, 1998, 2008). There are many studies stating that the Gaussian kernel is the unique kernel for generating a scale-space (Koenderink, 1984; Babaud *et al.*, 1986; Yuille & Poggio, 1986; Lindeberg, 1994). Convolving an image with Gaussian kernels and Gaussian derivatives has been regarded as an ensemble of image operators for the scale-space representation concept, which can be used to address a large variety of computer vision tasks, such as interest point detection, invariant and distinctive feature extraction, motion tracking, local image structure detection, image-based classification, and object recognition (Lindeberg, 1994; Kadir & Brady, 2001; Lowe, 2004; Mikolajczyk & Schmid, 2004; Bay *et al.*, 2006).

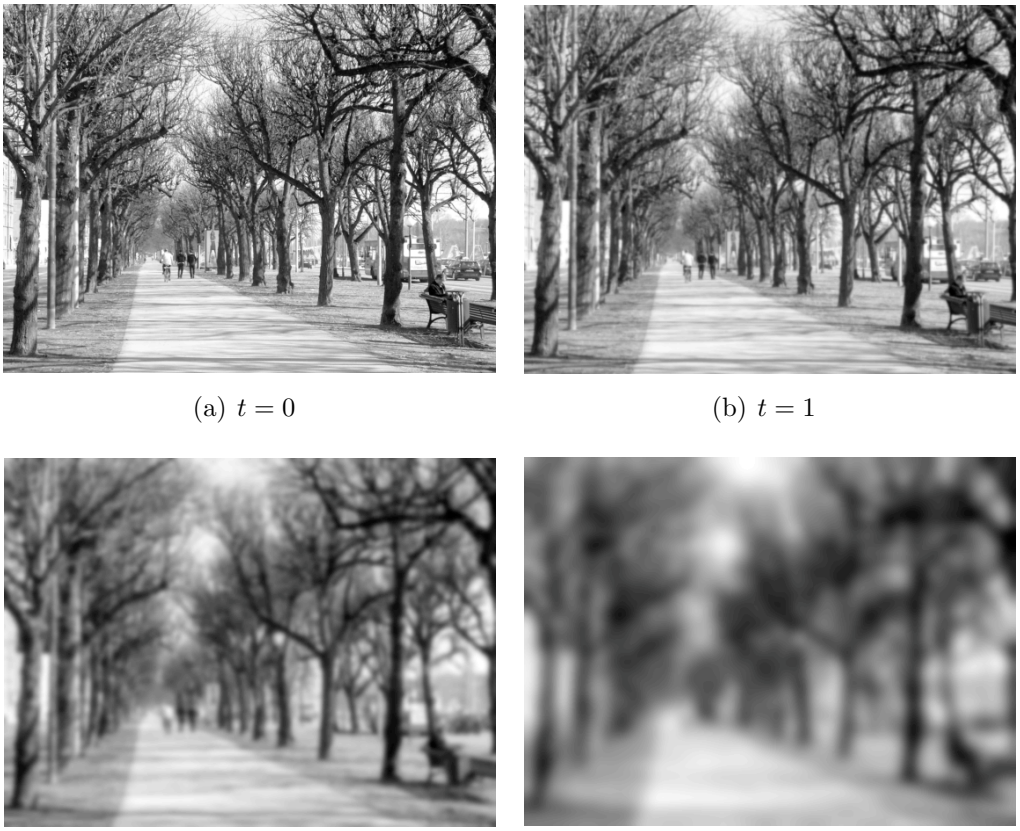
For any two-dimensional signal $f : \mathbb{R}^2 \to \mathbb{R}$, its scale-space representation $L : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ is defined by (Koenderink, 1984; Lindeberg, 1994, 1998, 2008):

$$L(x, y; t) = \int_{(\xi,\eta)\in\mathbb{R}^2} f(x - \xi, y - \eta) g(\xi, \eta; t) d\xi d\eta, \tag{5.1}$$

where $g : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ is the Gaussian kernel:

$$g(x, y; t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}, \tag{5.2}$$

where $t = \sigma^2$ is the variance of the kernel, which is the scale parameter in the scale-space family. For $t = 0$, g becomes an impulse function and $L(x, y; 0) = f(x, y)$. Therefore, the scale-space representation of an image at scale level $t = 0$ is the original image itself. As t increases, the width of Gaussian kernels becomes larger



(a) $t = 0$ (b) $t = 1$

(c) $t = 8$ (d) $t = 64$

Figure 5.1: Scale-space representations of an example greylevel image (560×420 pixels) at scale levels $t = 0$, 1, 8 and 64 (Lindeberg, 2008).

and larger, and more and more details are removed from the resulting L, where fine-scale structures of spatial size significantly smaller than $\sqrt{t}$ are smoothed at scale level t. The scale-space representations of an example image are shown in Figure 5.1. As shown in the figure, the tree tops are initially smoothed at scale level $t = 1$, the tree branches are further smoothed at scale level $t = 8$, and all details of the image are mixed at scale level $t = 64$.

Based on the scale-space representation, scale-space derivatives at any scale t can be computed in two ways. The first is differentiating the scale-space representation directly (Equation 5.3), and the second is convolving the original image with derivatives of Gaussian kernels (Equation 5.4).

$$L_{x^\alpha y^\beta}(x, y; t) = \partial_{x^\alpha y^\beta} L(x, y; t), \tag{5.3}$$

$$L_{x^\alpha y^\beta}(x, y; t) = (\partial_{x^\alpha y^\beta} g(x, y; t)) * f(x, y), \tag{5.4}$$

where $\alpha + \beta$ is the order of scale-space derivatives. These derivatives together constitute the basis for extracting features within the scale-space framework and provide a concise characterisation of local image structures. A second-order family

of scale-space derivatives include $\{L_x, L_y, L_{xx}, L_{xy}, L_{yy}\}$, which have been widely used for edge detection, blob detection, corner detection, and ridge detection (Lindeberg, 1998, 2008). The blob detection methods based on the scale-space framework will be reviewed in the following section.

5.2 Blob Detection

Many approaches have been developed so far for blob detection. Jackway (2000) proposed an improved morphological top-hat transform to detect light or dark blobs in a noisy image. Liu *et al.* (2010) defined a blobness measure based on eigenvalues of the Hessian matrix. Pixels with higher blobness values were labelled as blobs. There are also a number of approaches to detecting circles based on the Hough transform (Chiu & Liaw, 2005). In this section, we briefly review a set of blob detection methods which are based on scale-space analysis.

5.2.1 Laplacian of Gaussian

One of the most common blob detectors is based on the Laplacian of Gaussian (LoG) operator. Blobs can be detected at scale-space extrema of the normalised Laplacian $\nabla^2_{norm} L$, which are simultaneously local extrema with respect to both the spatial location and the scale level (Lindeberg, 1998). The normalised Laplacian is defined as:

$$\nabla^2_{norm} L = \sigma^2 (L_{xx} + L_{yy}), \tag{5.5}$$

where L is the scale-space representation of an image $I(x, y)$, produced from the convolution with a Gaussian kernel of various scales σ:

$$L(x, y; \sigma) = G(x, y; \sigma) * I(x, y), \tag{5.6}$$

where $*$ is the convolution operation and G is given by:

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{5.7}$$

The scale levels at which the extrema over scales are detected reflect the characteristic size of the blobs.

5.2.2 Difference of Gaussian

Lowe (2004) used the difference of Gaussian function (DoG) as a close approximation of the scale-normalised Laplacian of Gaussian, $\sigma^2 \nabla^2 G$, as studied by Lin-

deberg (1998):

$$G(x, y; k\sigma) - G(x, y; \sigma) \approx (k-1)\sigma^2 \nabla^2 G, \tag{5.8}$$

where k is a constant multiplicative factor between adjacent scales. Therefore, the detection of scale-space extrema of an image $I(x, y)$ can be achieved by detecting extrema in the difference of Gaussian function convolved with the image, denoted by $D(x, y; \sigma)$, which is computed by:

$$D(x, y; \sigma) = (G(x, y; k\sigma) - G(x, y; \sigma)) * I(x, y) = L(x, y; k\sigma) - L(x, y; \sigma), \tag{5.9}$$

where L is the scale-space representation of the image I, computed in the same way as Equation 5.6. It indicates $D(x, y; \sigma)$ is an efficient function to compute compared to the normalised Laplacian of Gaussian (Equation 5.5), as it can be simply computed by image subtraction.
There is a common drawback with the LoG and the DoG blob detectors, which is local extrema can also be present near contours or straight edges in an image, where the intensity has a relatively distinct change in one spatial direction. This results in the localisation of blobs is very sensitive to noise or small textures in the local neighbourhood.

5.2.3 Determinant of the Hessian Matrix

The Hessian matrix of a scale-space representation $L(x, y; \sigma)$ is defined as the matrix of second order partial derivatives:

$$H = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{yx} & L_{yy} \end{bmatrix}. \tag{5.10}$$

The trace of the scale-normalised Hessian matrix of L is defined as:

$$trace\ H_{norm}L = \sigma^2(L_{xx} + L_{yy}). \tag{5.11}$$

As can be seen, it is the scale-normalised Laplacian given by Equation 5.5. As stated above, blobs can be detected at scale-space extrema of the normalised Laplacian $\nabla^2_{norm}L$. Based on the studies in (Lindeberg, 1998), blobs can also be detected at scale-space extrema of the determinant of the normalised Hessian matrix of L:

$$det\ H_{norm}L = \sigma^2(L_{xx}L_{yy} - L^2_{xy}). \tag{5.12}$$

The location of blobs is determined by spatial extrema of the function of $detH_{norm}L$. The characteristic scale of blobs is automatically selected by searching for extrema over scales. The radius of the blobs is proportional to the selected scale. Detect-

ing blobs based on the determinant of the Hessian matrix can eliminate points at which the intensity changes in only one direction. This can effectively reduce the effects caused by contours or straight edges in the image.

5.2.4 Hessian-Laplacian

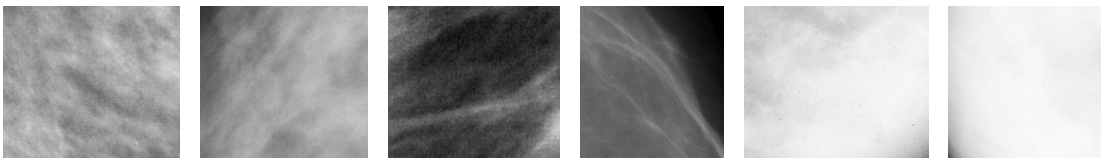
Mikolajczyk & Schmid (2004) proposed a hybrid method for blob detection by combining the trace and the determinant of the Hessian matrix of L, which is referred to as Hessian-Laplacian. The spatial location of blobs is determined by finding extrema of the determinant of the scale-normalised Hessian matrix of L, $det\, H_{norm}L$ (Equation 5.12), and automatic scale selection over scales is performed by searching for extrema of the trace of the scale-normalised Hessian matrix of L, $trace\ H_{norm}L$ (Equation 5.11), i.e. the scale-normalised Laplacian, $\nabla^2_{norm}L$ (Equation 5.5), at locations of spatial extrema.

5.2.5 Fast-Hessian

Bay *et al.* (2006) proposed a fast interest point detector, referred to as Fast-Hessian. The detector is based on the Hessian matrix. However, both the location and the scale are selected by relying on the determinant of the Hessian matrix rather than using the determinant and the trace of the Hessian matrix, respectively (as was performed in the Hessian-Laplacian detector (Mikolajczyk & Schmid, 2004)). The Gaussian second order partial derivatives are approximated with box filters. Due to the use of box filters and integral images, instead of iteratively applying the same filter to the down-sampled image in an image pyramid, filters of any size can be applied to directly filter the original image at exactly the same computational speed.

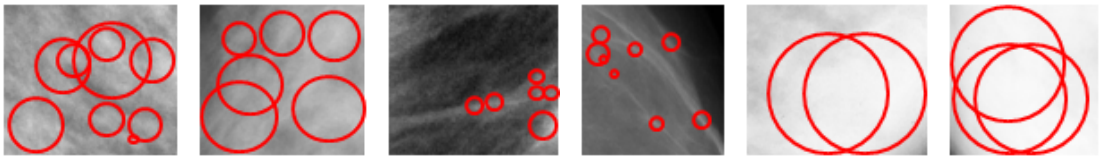
5.2.6 Salient Region

Kadir & Brady (2001) proposed a multiscale algorithm for the detection of salient regions of an image with automatic scale selection. A saliency metric is defined incorporating spatial position, feature space and scale simultaneously. This enables a comparison of the saliency of different features occurring at different spatial locations and scales. At each scale level, for each pixel location, the local entropy is calculated within a window of the size equal to the current scale. Afterwards, scales are selected at which peaks of the entropy are attained. The points with a saliency value above a threshold are measured as salient points and their surrounding areas are detected as salient regions. The selected scales at these points correspond to the sizes of the salient regions. Moreover, it has been pointed out



(a) (b) (c) (d) (e) (f)

Figure 5.2: Example breast tissue patterns: (a), (b) nodular tissue; (c), (d) linear tissue; (e), (f) homogeneous tissue.



(a) (b) (c) (d) (e) (f)

Figure 5.3: Pseudo blobs superimposed on image patches of example breast tissue patterns: (a), (b) nodular tissue; (c), (d) linear tissue; (e), (f) homogeneous tissue.

that this method favours blob-like image features and salient regions are more likely to be detected at blob-like structures.

5.3 A Blob Based Representation of Mammographic Parenchymal Patterns

Mammographic parenchymal patterns play an important role in mammographic risk assessment (see Section 1.3). Parenchymal patterns are determined by the spatial distribution of relatively dense tissue in the breast. As described by Tabár *et al.* (2005), there are four building blocks: nodular, linear, homogeneous and radiolucent. Effectively, we attempt to model nodular and homogeneous tissue in Tabár's model for mammographic risk assessment, as they have relatively high density and dominant effects on the overall density of the breast. Some example breast tissue patterns are shown in Figure 5.2 (image patches were manually selected by an expert radiologist). We assume that nodular and homogeneous tissue can be approximated as blob-like objects of various sizes. Therefore, we propose to use a set of multiscale blobs to represent parenchymal patterns in mammograms. In Figure 5.3, pseudo blobs (manually drawn) are superimposed on the example image patches of Figure 5.2, to provide an intuitive impression that different sized blobs can be detected from different types of breast tissue patterns.

As described in Section 5.1, scale-space theory provides a framework for multiscale image representation, which allows to consider representations of an image at multiple scales simultaneously. Blob-like structures can be detected at local

extrema in the scale-space signature. The size of the blobs can be estimated as the scale at which the extrema are obtained (Lindeberg, 1998). Thus, the detection of multiscale blobs in mammographic images forms the basis for the proposed representation of parenchymal patterns.

For mammographic image analysis, numerous methods for characterising parenchymal patterns have been developed (see Section 2.3 for details). Some approaches are also based on scale-space techniques (such as features extracted using a bank of filters derived from Gaussian kernels). However, none of the previous publications have suggested modelling breast tissue by focusing on blob-like tissue patterns.

It should be noted that there are two main differences between our modelling of mammographic parenchymal patterns and existing approaches (e.g. Karssemeijer (1998); Oliver *et al.* (2008); Petroudi *et al.* (2003); Subashini *et al.* (2010)). We do not apply an implicit intensity normalisation of mammographic images to overcome brightness and contrast variations in the acquisition process. In addition, we do not assume that mammograms in the same density class necessarily have the same global texture appearance. In our developed approach to representing breast parenchymal patterns, blobs can be regarded as local salient features over scales which accommodate large illumination variations. Furthermore, the spatial distribution of resulting multiscale blobs within the breast associated with their spatial/qualitative relations can provide a description of the anatomical structure of breast tissue, which has strong links with breast cancer risk. On the other hand, the modelling process of the proposed approach can be analogised as the mammogram perception process by radiologists. They are initially sensitive to salient tissue patterns and then classify mammograms into different density classes according to the spatial distribution and topology of these salient aspects.

In addition, the proposed approach is supported by mammogram synthesis work. In Bakic *et al.* (2002a), synthetic mammograms were generated by simulating breast tissue with large and medium scale tissue structures comprising different sized tissue elements (i.e. shells and blobs). The formation of breast parenchymal patterns was modelled by the projection of these compartments. We aim to reverse this generation process, and decompose breast tissue into a set of blobs at different scales (inclusion of shell-like structures is seen as future work).

5.3.1 Detection of Multiscale Blobs

The approach used to detect multiscale blobs in mammograms is based on the Laplacian of Gaussian (LoG) operator. As described in Section 5.2.1, blob detection is achieved by finding the extrema of the normalised Laplacian $\nabla^2_{norm} L = \sigma^2(L_{xx} + L_{yy})$ with respect to both space and scale.

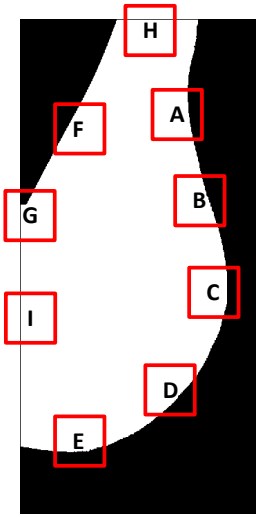


Figure 5.4: Example binary mask of breast region. Red squares indicate positions where the LoG filter goes outside the border of mask and/or image.

For our task of characterising parenchymal patterns, the region of interest in the image is solely the breast region. The background and the pectoral muscle are separated from the image using the method described in Section 3.2. An example binary mask of the breast region is shown in Figure 5.4. The separation of the background from the breast region produces a set of edges along the breast-background boundary (e.g. positions A, B, C, D and E in Figure 5.4), and the removal of the pectoral muscle results in a very sharp edge in the top right or left corner of the image (e.g. positions F and G in Figure 5.4). When filtering the image with the LoG operator, these false edges can interfere with blob detection, as the Laplacian has strong responses not only to blobs but also to edges. In addition, the image border can also affect the filtering, as the filter goes beyond the bounds of the image when convolving local windows around pixels near the border (e.g. positions G, H and I in Figure 5.4). In order to eliminate the effects caused by the boundary of the breast region and the border of the image, we propose a modification of the standard LoG filter to enable the spatial filtering adaptive in the sense that the filter template varies with the current processing window.
To illustrate the difference between the standard and modified LoG filters, we first generate a standard LoG filter denoted by $f_s(x, y)$:

$$f_s(x, y) = f(x, y) - \sum_{(x,y)\in T} f(x, y)/N_T, \tag{5.13}$$

where $f(x, y)$ is the LoG operator, T is the filter template which is a square local window, and N_T represents the number of pixels within the filter template. Therefore, the standard LoG filter is the normalised LoG operator where

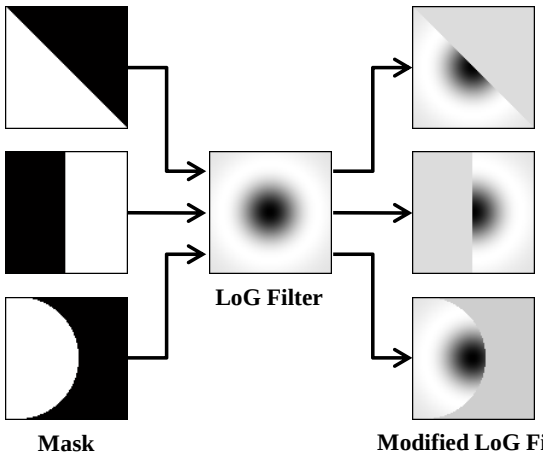


Figure 5.5: The standard LoG filter (centre) and three example modified LoG filters resulting from three different masks. The contrast of all filters has been normalised for better visualisation.

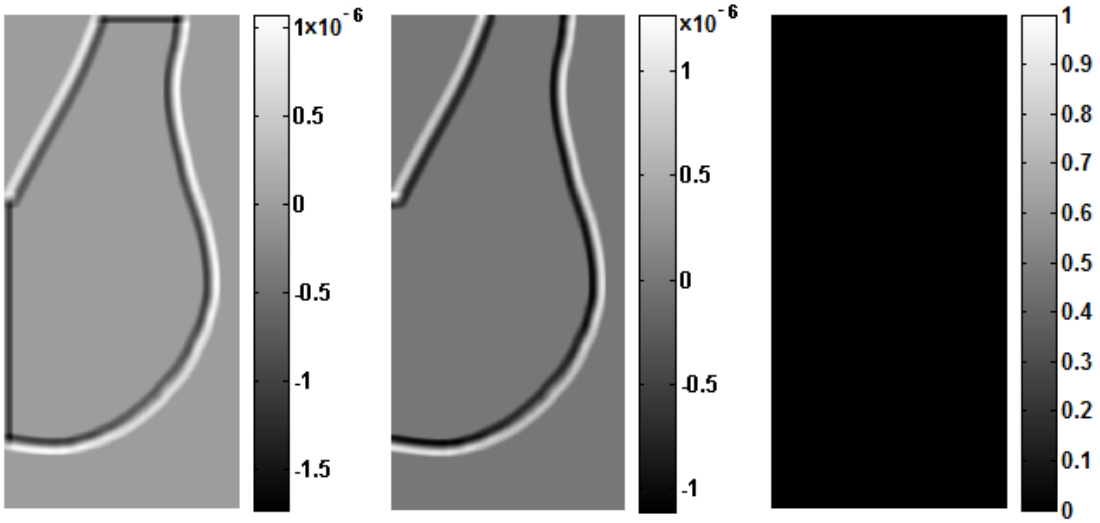
the sum of all coefficients within the filter template is normalised to zero, i.e. $\sum_{(x,y)\in T} f_s(x,y) = 0$. An example standard LoG filter is shown in the centre of Figure 5.5. The size of the filter is 101×101 pixels with standard deviation σ equal to 20 pixels. The purpose of normalising the LoG filter is to produce a very weak filter response to homogeneous regions (ideally equal to zero for purely uniform regions of constant pixel values).

For the modified LoG filter denoted by $f_m(x,y)$, we propose to apply a deformable filter template which can alter adaptively according to the local window to be convolved with it. Effectively, we define a mask (denoted by $M(x,y)$) to restrict the filtering to non-zero pixels within the local window under convolution: $\forall (x,y) \in T$, if pixel $I(x,y) \neq 0$, $M(x,y) = 1$; else $M(x,y) = 0$. The normalisation of the modified LoG filter is based on the mask area:

$$f_m(x,y) = \begin{cases} 0 & \text{if } M(x,y) = 0, \\ f(x,y) - \sum_{(x,y)\in T \& M(x,y)=1} f(x,y)/N_M & \text{if } M(x,y) = 1, \end{cases} \tag{5.14}$$

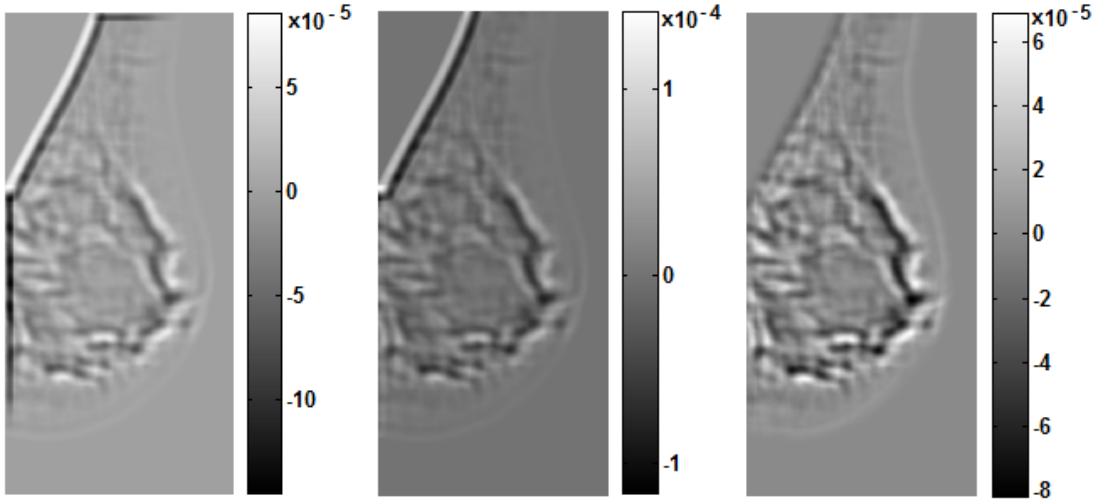
where $f(x,y)$ is the LoG operator and N_M is the number of non-zero pixels within the local window under convolution, computed by $N_M = \sum_{(x,y)\in T} M(x,y)$. As a result, the sum of the filter coefficients within the mask area is normalised to zero, i.e. $\sum_{(x,y)\in T \& M(x,y)=1} f_m(x,y) = 0$. Three example modified LoG filters corresponding to three masks are shown in Figure 5.5.

To demonstrate the modified LoG filter can effectively eliminate responses to the false edges, a direct comparison is made by filtering the binary mask of the breast region with the standard LoG filter and the modified LoG filter, respectively. For



(a) (b) (c)

Figure 5.6: Filtering results of the binary breast mask: (a) filtered image of the standard LoG filter using 0 padding; (b) filtered image of the standard LoG filter using replicate padding; (c) filtered image of the modified LoG filter.



(a) (b) (c)

Figure 5.7: Filtering results of the mammographic image: (a) filtered image of the standard LoG filter using 0 padding; (b) filtered image of the standard LoG filter using replicate padding; (c) filtered image of the modified LoG filter.

the standard LoG filter, two methods are adopted when the filter goes beyond the image border. One option is padding the outside part with 0s (0 padding), the other is assuming the pixels outside the image border equal to the nearest border pixel (replicate padding). The filtered images are shown in Figure 5.6. As can be seen from the filtered images of the standard LoG filter, positive and negative filter responses appear along the boundary of the breast mask (Figure 5.6(a) and 5.6(b)) and negative filter responses are present near the border of

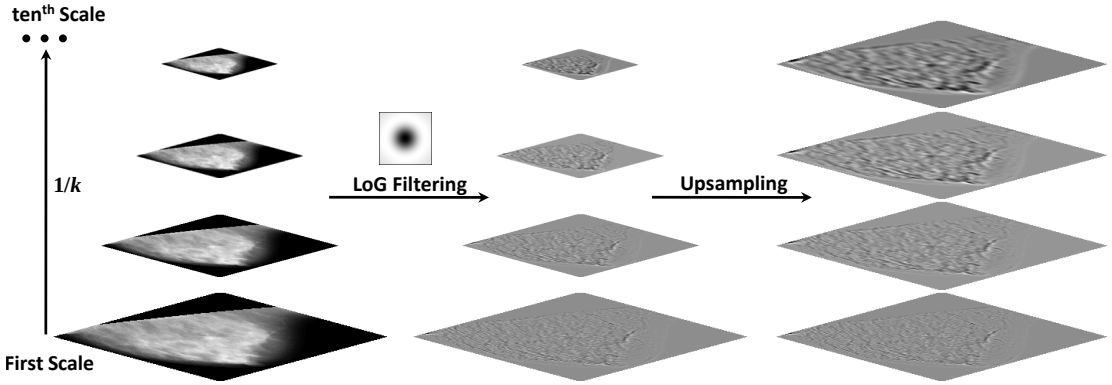
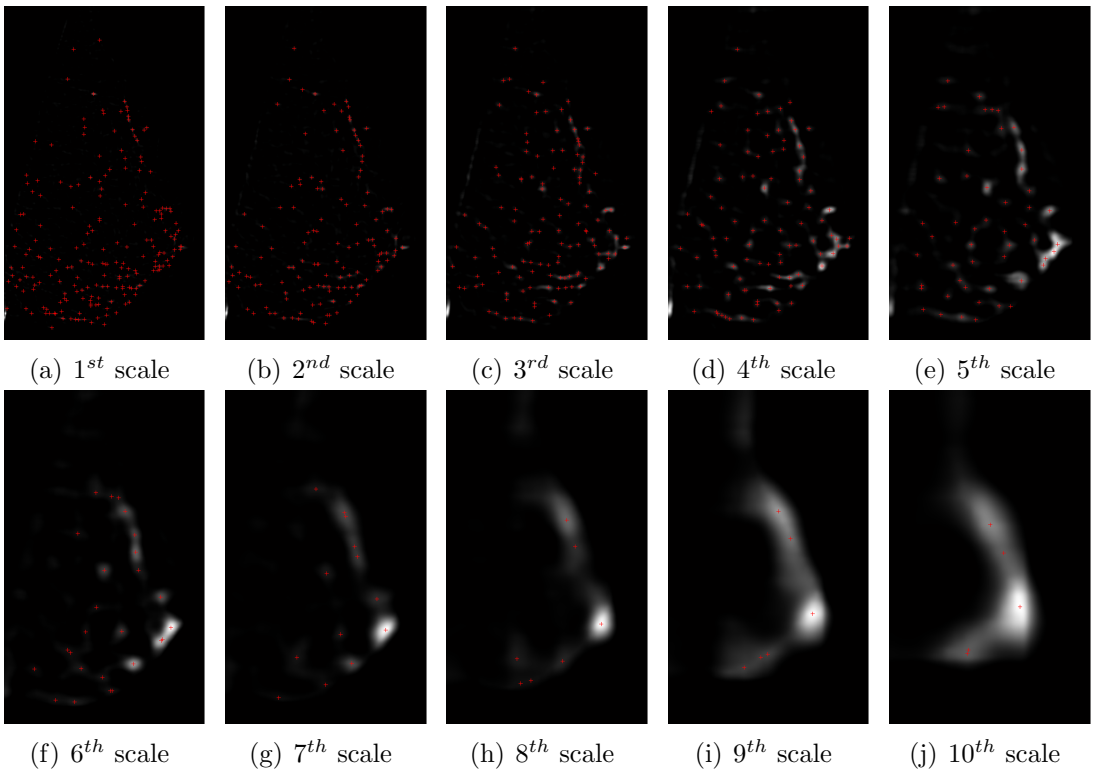


Figure 5.8: Multiscale LoG filtering. The initial image is down-sampled by a factor $1/k$ and incrementally filtered with one LoG filter. All filtered images are then upsampled to the initial size.

the image when using 0 padding (Figure 5.6(a)), which will affect the accuracy of the blob detection. In contrast, the modified LoG filter indicates zero responses to the uniform breast mask. A further demonstration is provided by filtering the mammographic image from which the binary mask (Figure 5.4) is extracted. The filtered images of the corresponding mammographic image are provided in Figure 5.7. It is shown that the effects caused by the boundary of the pectoral muscle (Figure 5.7(a) and 5.7(b)) and the image border (Figure 5.7(a)) are significantly reduced by the modified LoG filter (Figure 5.6(c)).

We detect blobs from a range of ten scales, where the first scale $\sigma_1 = 8$ and the scale factor $k = \sqrt{2}$. For efficiency reasons, we avoid repeatedly filtering the mammographic image (which normally has a large size due to the high spatial resolution of mammography) with a filter of increasing size across the scale space (the size of the filter has a linear relationship with the scale). We down-sample the image by a factor $1/k$ instead of increasing the filter by the factor k at each scale. We then upsample all filtered images in the hierarchical construction to the full size for searching for the extrema over scales. The procedure of multiscale LoG filtering is shown in Figure 5.8. The tissue patterns of interest here are high-density regions, which are detected in the form of bright blobs (indicated by negative filter responses), so positive filter responses in filtered images are removed (set to zero). The squared LoG filter responses of an example mammogram at ten scales are shown in Figure 5.9 where positive responses have been removed.

To search for the scale-space extrema, we first detect extrema from each scale separately. Each pixel located in the breast region is compared with its neighbours within a local region in the current scale-space image. We use 5×5 local regions, so each pixel is compared with its 24 neighbours at each scale. The extrema with low contrast (less than a threshold) are discarded. The threshold can be dynamically



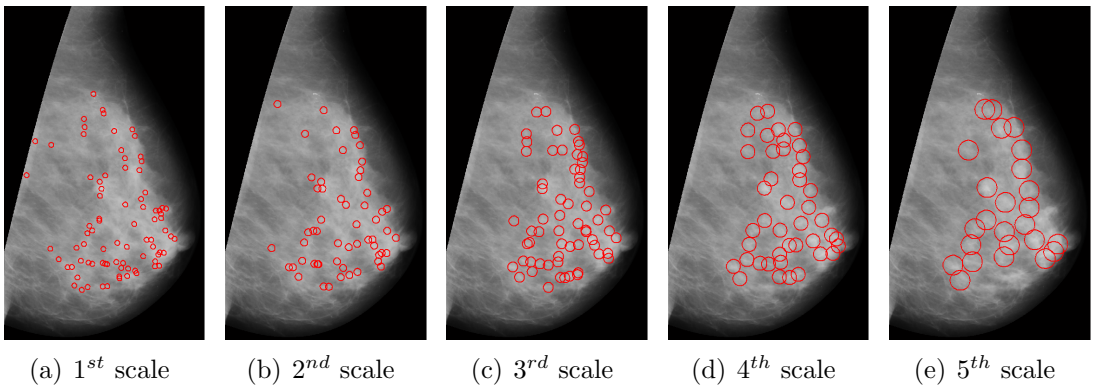
(a) 1^{st} scale (b) 2^{nd} scale (c) 3^{rd} scale (d) 4^{th} scale (e) 5^{th} scale

(f) 6^{th} scale (g) 7^{th} scale (h) 8^{th} scale (i) 9^{th} scale (j) 10^{th} scale

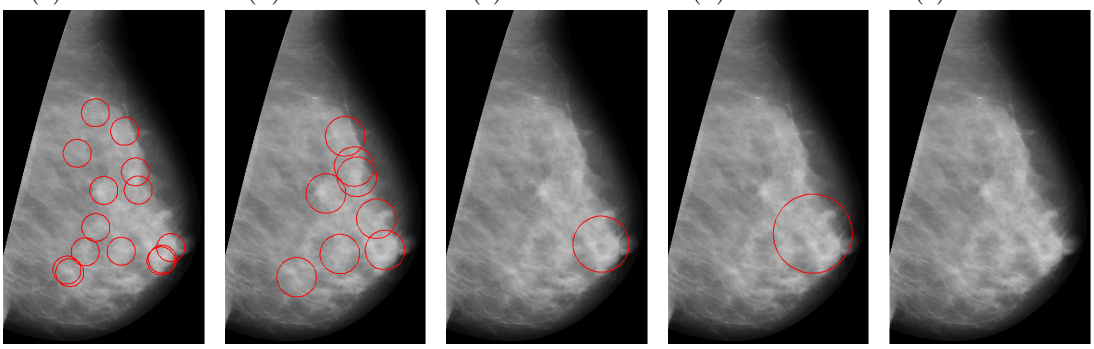
Figure 5.9: The squared LoG filter responses of an example mammogram at ten scales (positive responses have been removed) in which the obtained extrema after the initial thresholding based on contrast and spatial information are labelled with red pluses.

altered according to the maximal pixel value in the image. Every extremum in each scale-space image indicates the centre point of one blob, the radius of which is the corresponding scale. As described earlier, we aim to detect interior blob structures of the breast, and therefore blobs whose major covering regions appear outside the breast are ignored. This step can be regarded as contrast based thresholding taking spatial information into account. The locations of the resulting extrema are labelled with red pluses in Figure 5.9.

In addition, in order to capture true blob-like dense breast tissue, it is not sufficient simply to reject the extrema with weak responses. At smaller scales, some strong responses can be induced by linear structures (vessels in particular), as the LoG filter also has the capability of detecting edges. On the other hand, at larger scales, some false extrema/blobs (non-dense tissue) tend to appear along the breast boundary, due to the influence of the natural profile of the breast. The reduction of such false positives can be performed by means of putting a threshold on the mean intensity within the corresponding blob area. To calculate the threshold, the modified FCM algorithm described in Section 3.3 is used to partition breast tissue pixels in the initial image into nine clusters. The difference between the mean intensity and the standard deviation of clusters from six to nine (corresponding



(a) 1^{st} scale (b) 2^{nd} scale (c) 3^{rd} scale (d) 4^{th} scale (e) 5^{th} scale

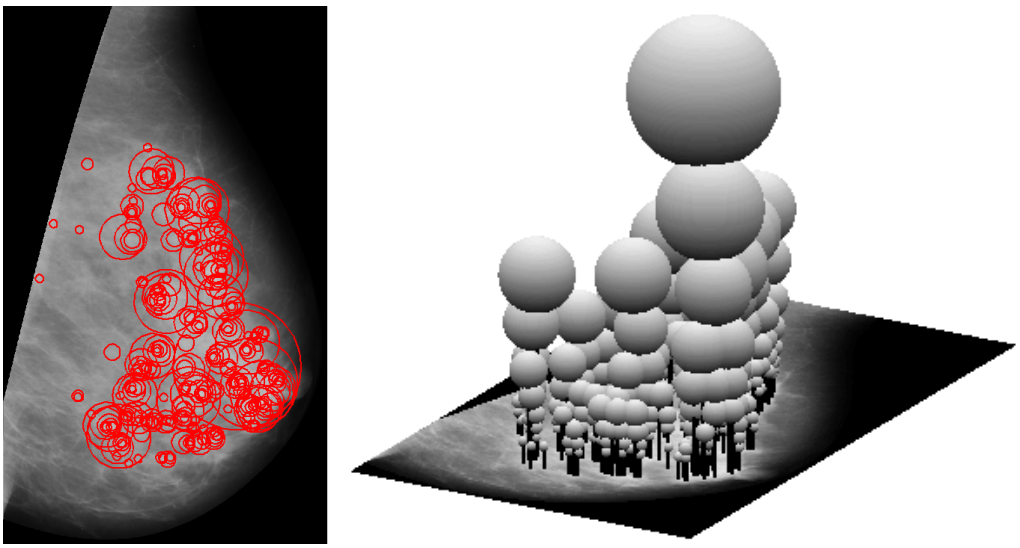


(f) 6^{th} scale (g) 7^{th} scale (h) 8^{th} scale (i) 9^{th} scale (j) 10^{th} scale

Figure 5.10: Detected blobs at each scale after removing false positives. The previous extrema whose corresponding blobs have lower mean intensity are regarded as false positives (corresponding to non-dense tissue) and removed (e.g. all extrema at the 10^{th} scale in Figure 5.9).

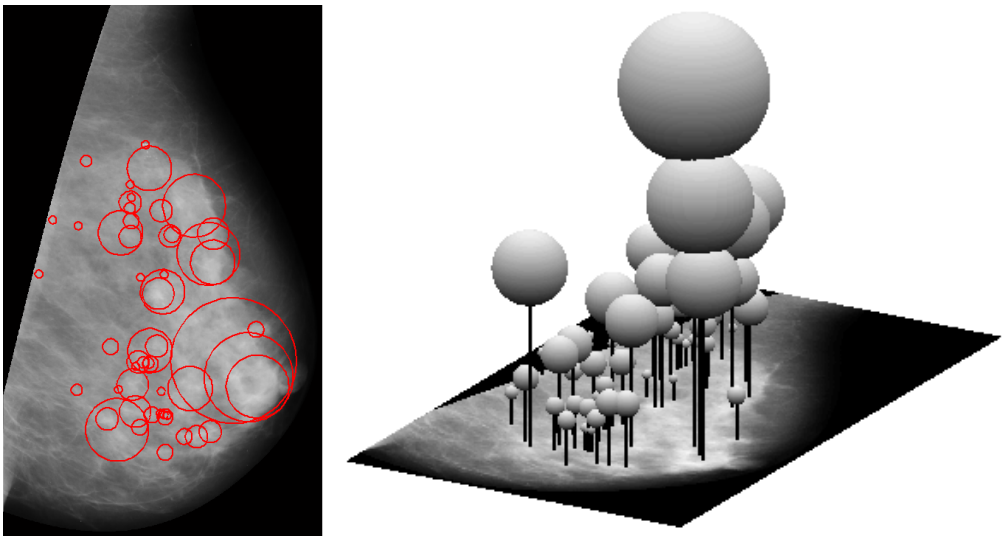
to dense tissue) is chosen as the threshold. Alternative number of clusters and those selected to represent dense tissue can provide similar results. The extrema whose corresponding blobs have lower mean intensity than the threshold are removed. Blobs detected at each scale are shown in Figure 5.10, where blobs are indicated with red circles and superimposed on the segmented breast region of the mammogram. All blobs detected from the ten scales are shown in Figure 5.11. A three-dimensional view is also provided where the mammographic image is shown in the horizontal plane, the vertical dimension corresponds to scale, and blobs are displayed as spheres in scale space.

After identifying the extrema from 5×5 local regions at each individual scale, we will select the global extrema from them over all scales. Each is compared with its $5 \times 5 \times 9$ "neighbours" in the other scale-space images. Finally, only the global extrema of $5 \times 5 \times 10$ blocks in the whole scale space are retained as blob candidates with both locations and scales identified. The resulting blobs are shown by both 2D and 3D views in Figure 5.12. As can be seen that concentric/overlapping blobs which are located at different scale planes are combined into the one with the maximum response.



(a) (b)

Figure 5.11: All blobs detected from the ten scales: (a) blobs shown as red circles in the mammographic image; (b) blobs shown as spheres in a three-dimensional space (the mammographic image is shown in the horizontal plane, and the vertical dimension corresponds to scale).

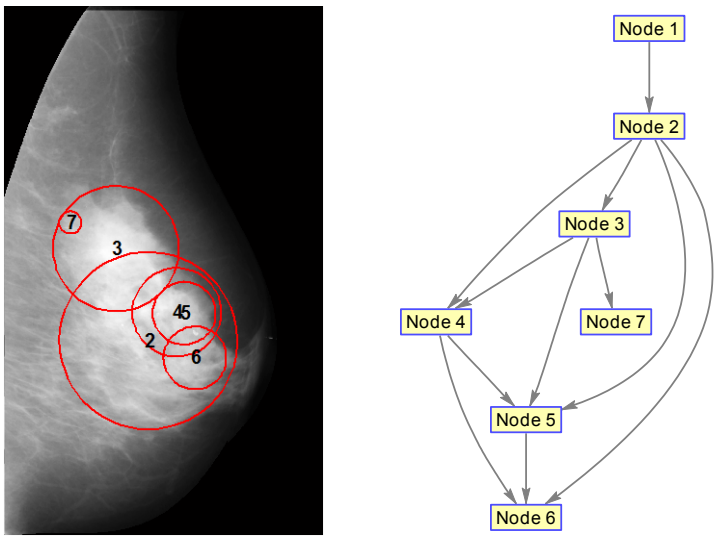


(a) (b)

Figure 5.12: Final detection results after combining blobs over scales: (a) blobs shown in 2D; (b) blobs shown in 3D.

5.3.2 Blob Merging

The blobs detected above mainly correspond to the dense tissue regions which have approximately blob-like structures. Blob overlapping happens among those which are closely located in dense tissue. To interpret an overall outline of dense tissue, we merge the overlapping blobs based on their qualitative relations. Here, we



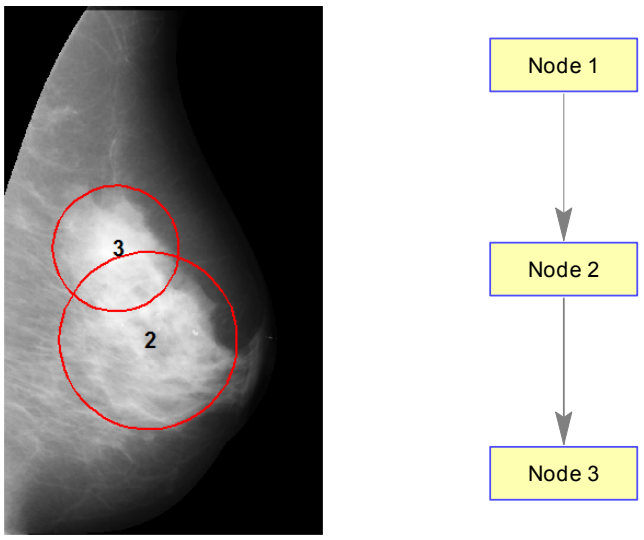
(a) (b)

Figure 5.13: A simple example of blob graph: (a) blobs ordered with a sequential number staring from 2; (b) blob graph.

consider three types of qualitative relations: external, intersection, and internal. The qualitative relation between two blobs (e.g. blob A and blob B) depends on their respective radii (denoted by r_A and r_B) and the distance (denoted by d) between their centre points. The decision conditions of the three relations are defined as (assuming $r_A \geq r_B$): $d \geq r_A + r_B$ (external), $r_A - r_B < d < r_A + r_B$ (intersection), and $d \leq r_A - r_B$ (internal), respectively.

To describe the qualitative relations among all the blobs, a directed blob graph is constructed, which can effectively describe the topology of the blobs. Each blob is labelled with a sequential number which is ordered according to the scale and position of the blob (from the coarse scale to the fine scale and in a left-to-right and top-to-bottom direction for each scale). In the blob graph, each node represents a blob, and each arrow indicates the two corresponding blobs overlap with each other (intersection or internal). The direction of the arrow is from the lower to the higher numbered node. A blob cluster comprise a subgraph where the blob with the smallest sequential number is chosen as the root node. A top root node (i.e. node 1, representing the whole breast) is used to form a single rooted graph. A simple example is shown in Figure 5.13. Here, the blob graph is only used for blob merging and is not involved in the subsequent step of blob encoding. The inclusion of graph based features for breast density classification might be investigated in future work.

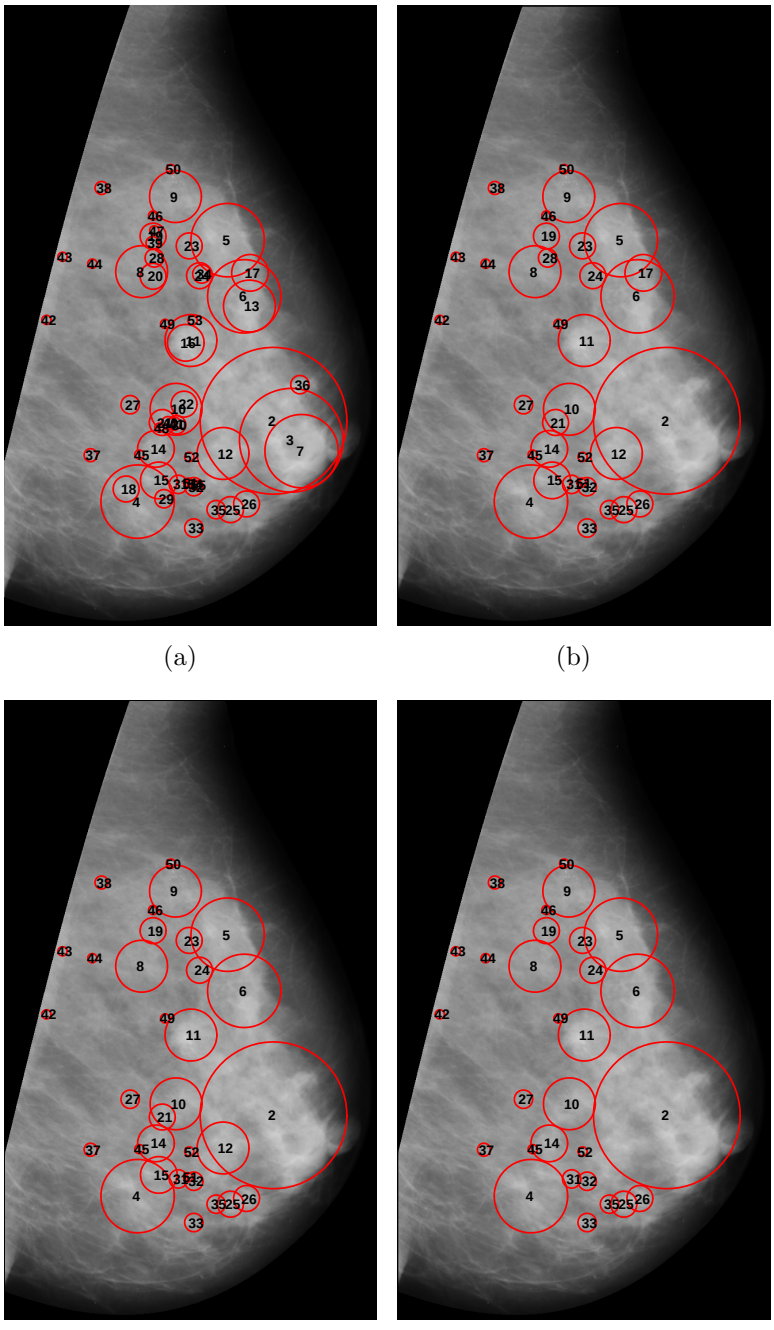
The merging procedure starts from the finest scale and proceeds to coarser scales, which is based on the following criteria. All external blobs separately located in the image will be retained, while internal blobs contained by larger ones will be



(a) (b)

Figure 5.14: Blob merging: (a) remaining blobs after blob merging; (b) new blob graph ($\alpha = 0.8$).

removed. When two blobs A (scale and radius denoted by σ_A and r_A) and B (scale and radius denoted by σ_B and r_B) intersect with each other, if they are closely located (assuming $r_A \geq r_B$, the centre point of blob B is located in blob A, i.e. $d \leq r_A - \alpha r_B$, $0 < \alpha < 1$), the integration of the Gaussian scale-space signature over the blob area will be calculated by $\int_{blob} g(x, y; \sigma^2) * I(x, y) dxdy$ and the blob with the larger value will be retained, otherwise both blobs will be retained. The remaining blobs after the blob merging and the new blob graph are shown in Figure 5.14. It is shown that the structure of the blob graph is simplified and the overall appearance of dense tissue within the breast is preserved by the remaining blobs. This indicates the obtained blobs can be regarded as an overall description of dense tissue in the breast and benefit the qualitative analysis of breast density. As described above, there is a parameter α in blob merging, which controls the overlapping extent of the blobs after merging. The blob merging results of Figure 5.12 using three values of α $(0.3, 0.6, 0.9)$ are shown in Figure 5.15. There are 54 blobs in the initial detection result (Figure 5.15(a)), which are ordered with sequential numbers from 2 to 55. When using $\alpha = 0.9$, blobs {3, 7, 13, 16, 18, 20, 22, 29, 30, 34, 36, 39, 40, 41, 47, 48, 53, 54, 55} are removed and the number of blobs is reduced to 35 (Figure 5.15(b)); when using $\alpha = 0.6$, blobs $\{17, 28\}$ are further removed from the previous result and 33 blobs are left (Figure 5.15(c)); and for $\alpha = 0.3$, four more blobs $\{12, 15, 21, 29\}$ are removed and the number of blobs is 29 (Figure 5.15(d)).



(a) (b) (c) (d)

Figure 5.15: Blob merging for different values of α: (a) initial 54 blobs numbered from 2 to 55; (b) 35 blobs are retained when $\alpha = 0.9$; (c) 33 blobs are retained when $\alpha = 0.6$; and (d) 29 blobs are retained when $\alpha = 0.3$.

5.3.3 Blob Encoding

In addition to the qualitative analysis of breast density, the blobs achieved above can be used to quantitatively analyse breast density. For quantitative analysis, we define a numerical measure of breast density by means of encoding the multiscale blobs. The distribution of the blobs over scales is represented by a vector $N = (N_1, N_2, \ldots, N_n)$, where N_i $(i = 1, 2, \ldots, n)$ is the number of blobs at scale i, and n is the number of scales. The vector N is transformed into a unique number by encoding its elements with:

$$BLOB = \sum_{i=1}^{n} N_i k^{2(i-1)} \tag{5.15}$$

where k is the scale factor (equal to $\sqrt{2}$ in our work). The resulting value of $BLOB$ is the defined measure of breast density, which can be used for breast density classification. To avoid bias caused by the breast size, it is normalised by:

$$BLOB_{norm} = (A_{blob}/A_{breast})BLOB \tag{5.16}$$

where A_{blob} denotes the area of the blob at the finest scale, which is computed by $A_{blob} = \pi\sigma_1^2$ (σ_1 is the first scale equal to 8 pixels), and A_{breast} represents the area of the breast, which is the number of pixels within the breast region. Note that the value of A_{blob} is the same for all mammograms and seems redundant, but we aim to obtain a dimensionless measure and therefore normalise $BLOB$ with the ratio of A_{blob} and A_{breast}. In addition, by normalising the density measure in this way, it can be deemed to be an approximation of breast percent density, which approximately presents the relative proportion of dense tissue in the whole breast.

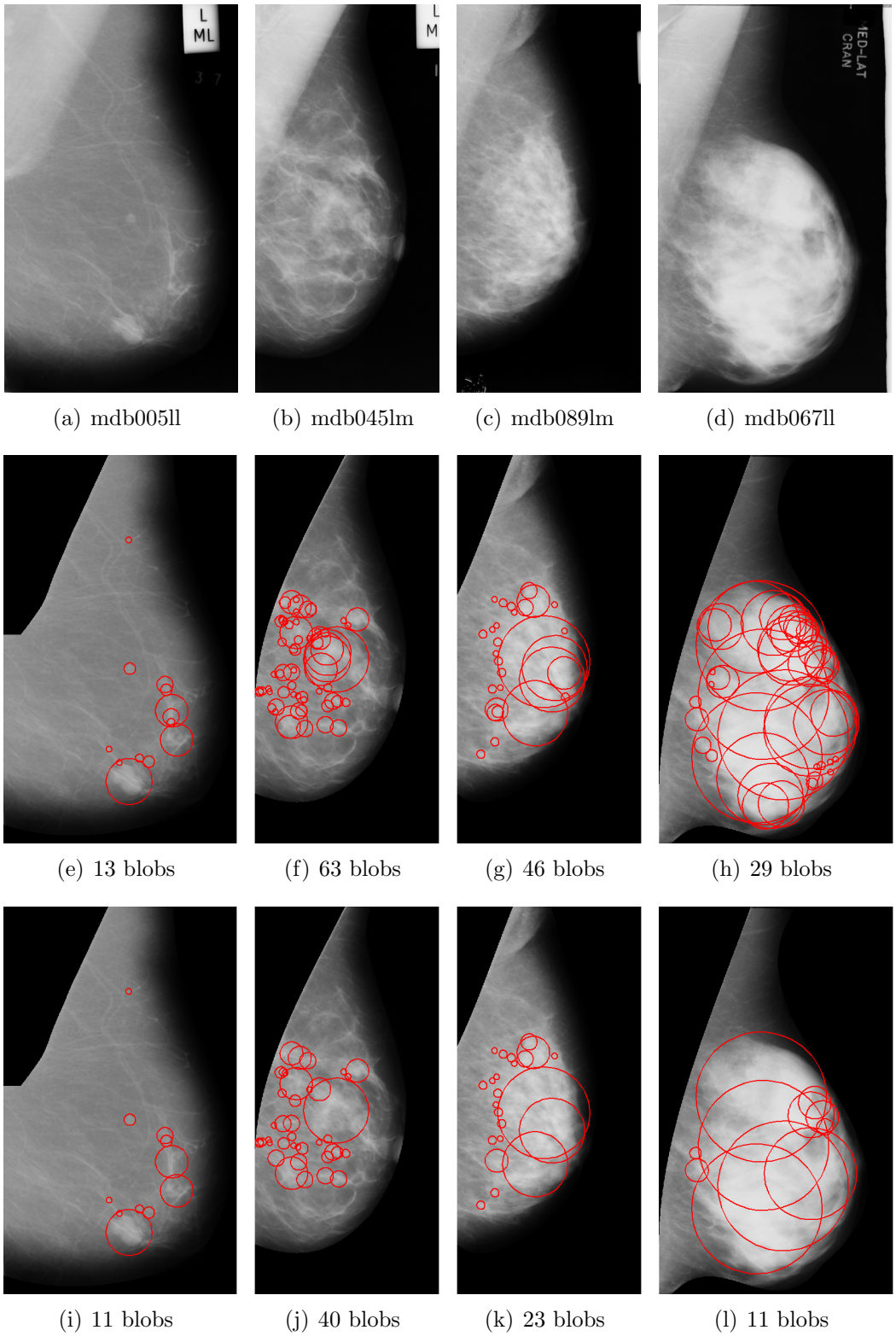
5.4 Results and Discussion

To evaluate the proposed blob based representation of mammographic parenchymal patterns, it has been tested using the full MIAS database (containing 321 left/right MLO mammograms) and a large dataset (containing 831 right MLO mammograms) from the DDSM database. As described above, blobs were detected at ten scales (starting from 8 pixels and increasing by a factor of $\sqrt{2}$).
Four example mammograms covering the four BIRADS classes were selected from MIAS and DDSM, respectively. The detection results of multiscale blobs are shown in Figures 5.16 and 5.17, where the top row shows the original mammograms, the middle row shows the results before blob merging, and the bottom row shows the results after blob merging ($\alpha = 0.8$). For BIRADS I, the breast is almost entirely fatty, only a few small scale blobs are detected; for BIRADS II, some small and

medium scale blobs are located at scattered dense tissue regions; for BIRADS III, a few relatively larger scale blobs appear in the dense tissue area; while for BIRADS IV, the breast is extremely dense and the whole breast is mainly covered by large scale blobs.

The corresponding three-dimensional views of the resulting blobs (after blob merging) are shown in Figures 5.18 and 5.19, where the distribution of these blobs over the ten scales is shown as bar charts under the 3D views. This indicates the number of blobs at larger scales increases with respect to the increasing BIRADS class. For the four MIAS examples, the density measure of "$BLOB$", computed by Equation 5.15, is 153, 378, 601 and 2460, and the corresponding normalised value "$BLOB_{norm}$", computed by Equation 5.16, is approximately equal to 0.0612, 0.2673, 0.4173 and 1.2701 ($A_{blob} \approx 200$ pixels, $A_{breast} = 499987$, 282844, 288076 and 387378 pixels). For the four DDSM examples, the density measure of "$BLOB$" is 9, 97, 341 and 470, and the corresponding normalised value "$BLOB_{norm}$" is approximately equal to 0.0111, 0.2113, 0.6775 and 0.8909 ($A_{blob} \approx 200$ pixels, $A_{breast} = 162225$, 91805, 100663 and 105507 pixels).

In our work, the Laplacian of Gaussian based blob detector was used for the detection of blobs at multiple scales, while other blob detectors described in Section 5.2 (e.g. the difference of Gaussian (Lowe, 2004), the determinant of the Hessian matrix (Lindeberg, 1998), Hessian-Laplace (Mikolajczyk & Schmid, 2004), and Fast-Hessian (Bay *et al.*, 2006)) could also be exploited. The scale range (from 8 to 181 pixels) was set empirically in our experiments, according to the size of mammographic images to be analysed, which can be tuned to specific applications. Moreover, we used a greylevel based threshold to reduce the false positives from the initial detection results. In future work, we will further investigate the aspects of both false positive and false negative reduction. At the blob merging step, some overlapping blobs were removed to simplify the topological structure of the blobs and then were fused into a general blob based representation of parenchymal patterns. The effects of this merging step will be further investigated. In addition, the tissue appearance within individual blob regions could be analysed. The texture signature of different sized blobs could be incorporated into the proposed representation which is currently only based on the blob distribution over scales. The defined measure of breast density $BLOB/BLOB_{norm}$ derived from the obtained blobs can be used for mammographic risk assessment. Further evaluation of the proposed method for mammographic risk assessment will be done in Chapter 6, based on the full MIAS database and the selected subset of the DDSM database.

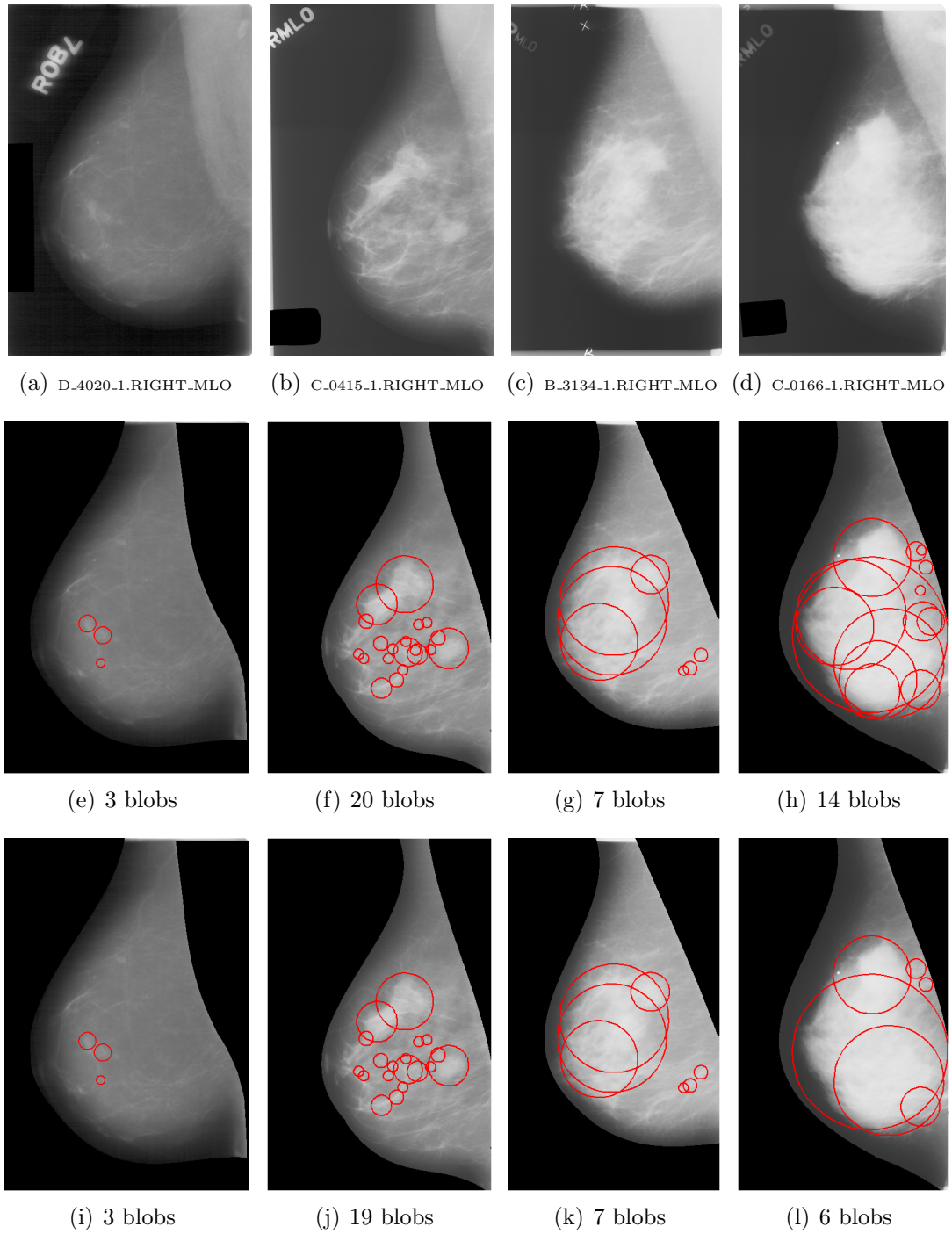


(a) mdb005ll (b) mdb045lm (c) mdb089lm (d) mdb067ll

(e) 13 blobs (f) 63 blobs (g) 46 blobs (h) 29 blobs

(i) 11 blobs (j) 40 blobs (k) 23 blobs (l) 11 blobs

Figure 5.16: Detection results of multiscale blobs for example mammograms from the MIAS database. Top row: original mammograms; middle row: resulting blobs before blob merging; bottom row: remaining blobs after blob merging. From left to right, the four mammograms are sorted from BIRADS I to BIRADS IV.



(a) D_4020_1.RIGHT_MLO (b) C_0415_1.RIGHT_MLO (c) B_3134_1.RIGHT_MLO (d) C_0166_1.RIGHT_MLO

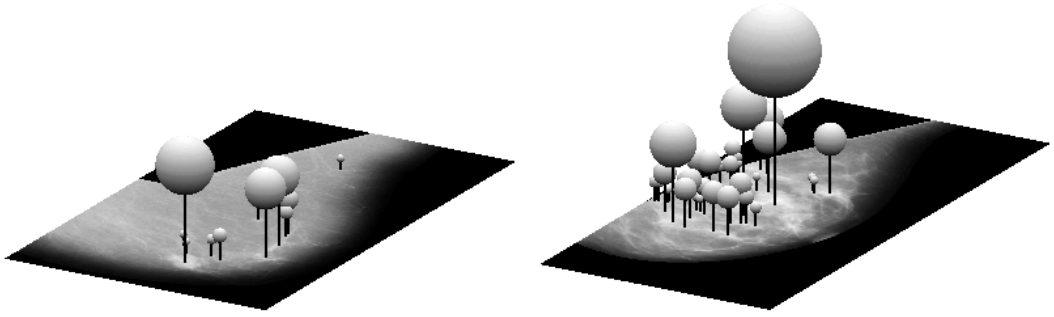
(e) 3 blobs (f) 20 blobs (g) 7 blobs (h) 14 blobs

(i) 3 blobs (j) 19 blobs (k) 7 blobs (l) 6 blobs

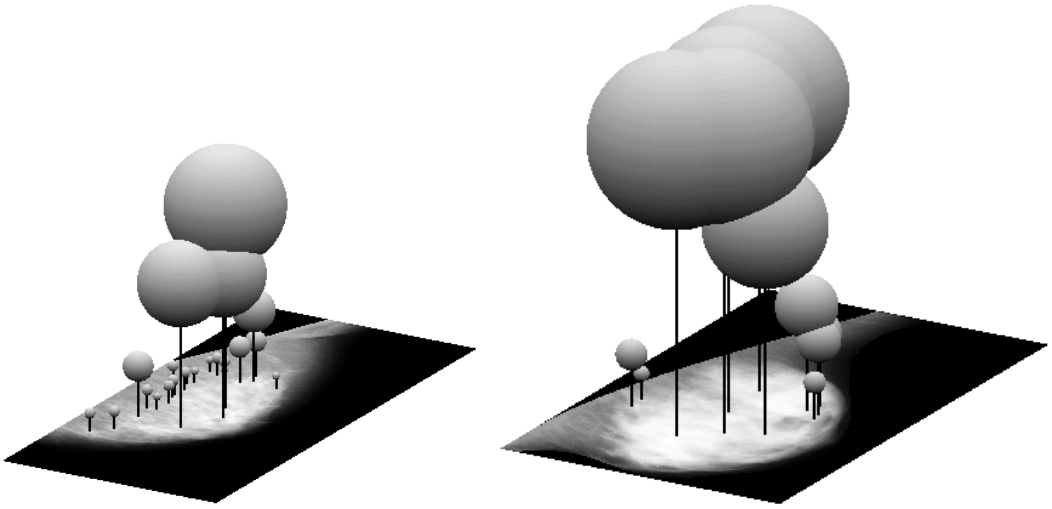
Figure 5.17: Detection results of multiscale blobs for example mammograms from the DDSM database. Top row: original mammograms; middle row: resulting blobs before blob merging; bottom row: remaining blobs after blob merging. From left to right, the four mammograms are sorted from BIRADS I to BIRADS IV.

5.5 Summary

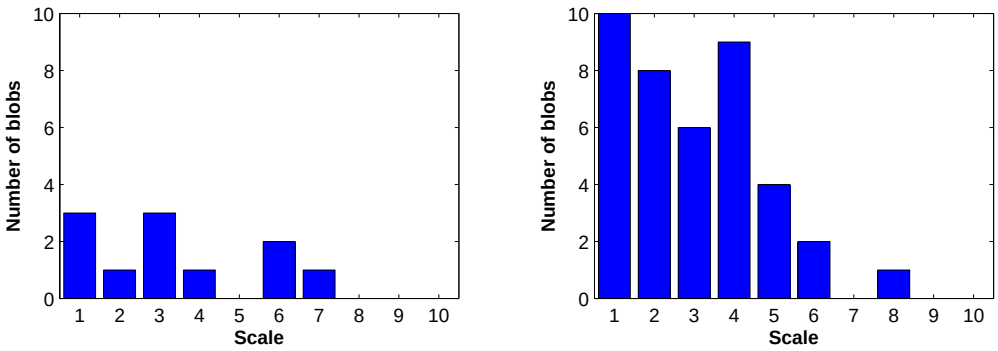
In this chapter, we investigated the detection of blob-like image structures based on multiscale image representation. At the beginning of the chapter, scale-space theory was introduced and a set of blob detectors were reviewed, including the Laplacian of Gaussian, the Difference of Gaussian, the determinant of the Hessian matrix, Hessian-Laplacian, Fast-Hessian, and salient region detection. On the basis of that, we proposed a multiscale blob based representation of mammographic parenchymal patterns. This idea was originated from the prior knowledge of the anatomical structure and the synthetic elements of breast tissue, obtained from previous publications (Tabár *et al.*, 2005; Bakic *et al.*, 2002a). For mammographic images, we detected blob-like regions by searching for local extrema in scale space, and used multiscale blobs to represent the approximately blob-like tissue patterns. The qualitative relations and the overall distribution of these blobs with respect to scales were combined to define a density measure of the breast. To our knowledge, the proposed method is a first attempt to model parenchymal patterns and analyse breast density only based on the dense regions with blob-like structures. It has been tested using the full MIAS database and a large subset of the DDSM database (see Chapter 6). The proposed blob based representations have indicated discriminative models of different parenchymal patterns, allowing to understand breast tissue patterns from a different viewpoint. In addition, as shown by the obtained results, the defined density measures are consistent with BIRADS density classes. Further evaluation for mammographic risk assessment will be provided in Chapter 6.



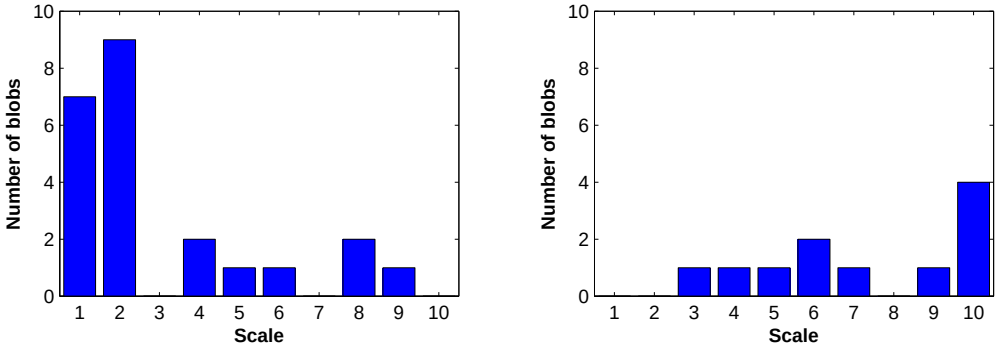
(a) BIRADS I (b) BIRADS II



(c) BIRADS III (d) BIRADS IV

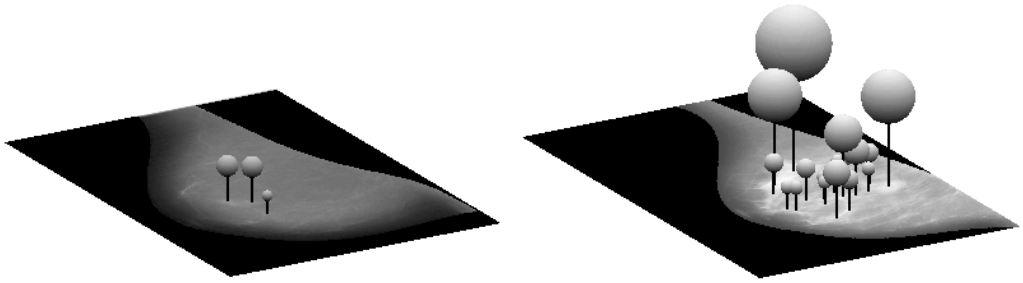


(e) BIRADS I (f) BIRADS II

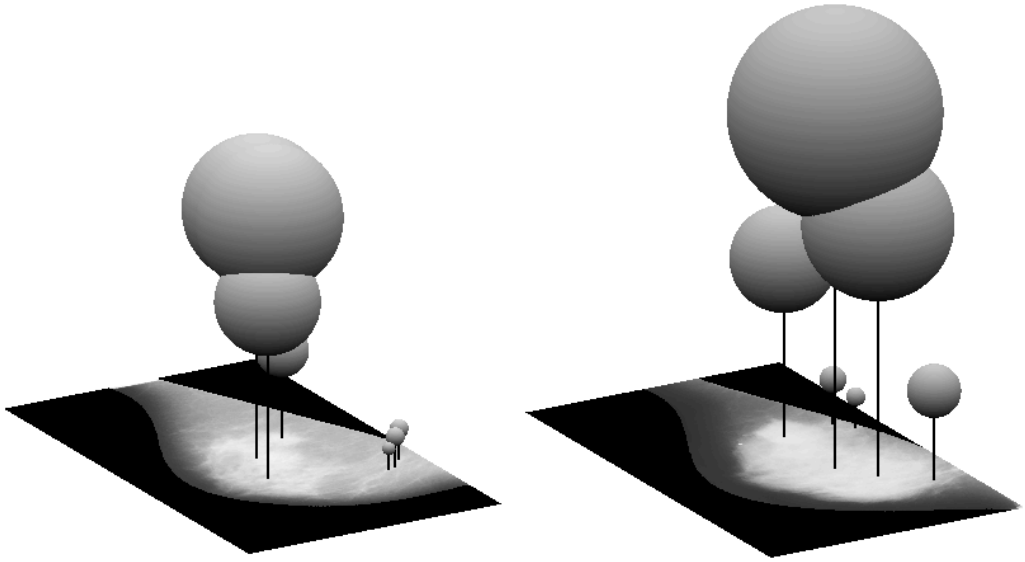


(g) BIRADS III (h) BIRADS IV

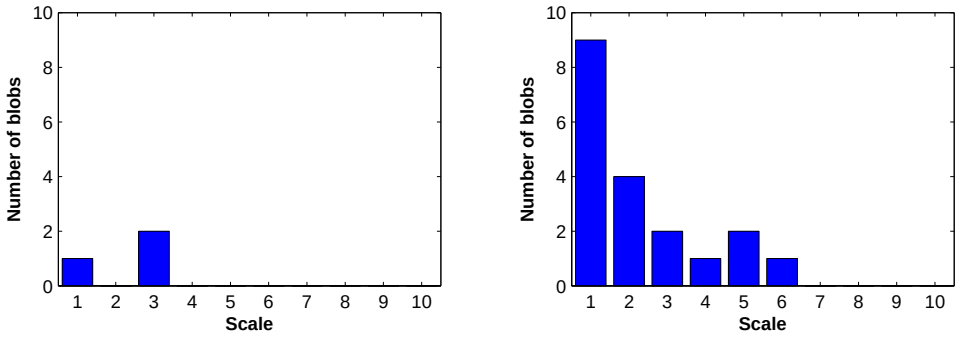
Figure 5.18: (a) - (d) Three-dimensional views of the resulting blobs for the four MIAS examples; (e) - (h) The distribution of these blobs over the ten scales.



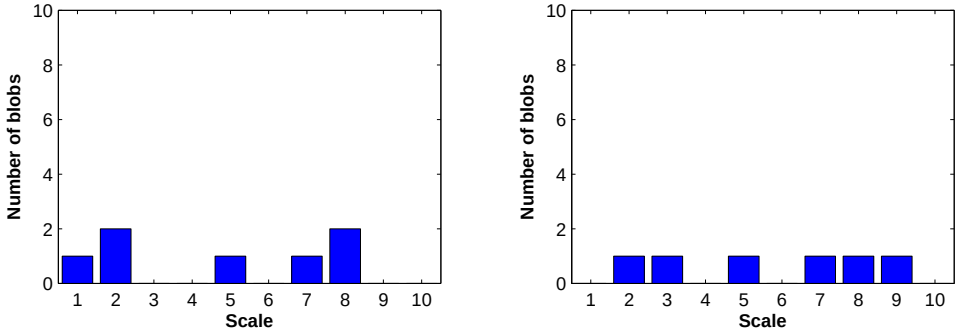
(a) BIRADS I (b) BIRADS II



(c) BIRADS III (d) BIRADS IV



(e) BIRADS I (f) BIRADS II



(g) BIRADS III (h) BIRADS IV

Figure 5.19: (a) - (d) Three-dimensional views of the resulting blobs for the four DDSM examples; (e) - (h) The distribution of these blobs over the ten scales.

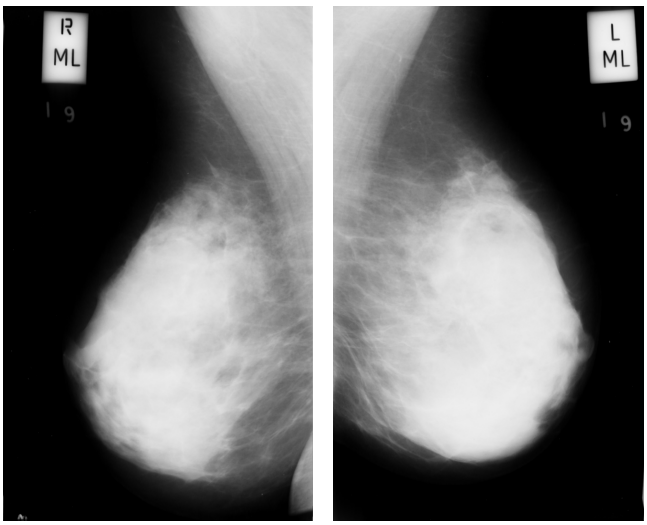
Chapter 6

Mammographic Risk Assessment

In the previous chapters, we have presented various methods for mammographic image analysis. In Chapter 3, we proposed three methods for the segmentation of mammographic images. The first method is regarding breast region segmentation, which is a pre-processing step to segment the breast region from the background and remove the pectoral muscle (if present). The other two methods are for breast density segmentation, which are based on a modified fuzzy c-means algorithm and a topographic representation, respectively. One segments the breast region into a number of sub-regions with different densities (e.g. fatty, semi-fatty, semi-dense, and dense tissue), and the other detects only dense tissue regions from the breast. In Chapter 4, we investigated five strategies for using different types of local features to characterise breast tissue appearance in mammographic images. Mammographic tissue appearance is modelled with occurrence histograms based on statistical analysis of local tissue appearance. In Chapter 5, we proposed a multiscale blob based representation of mammographic parenchymal patterns where blob-like dense regions construct the basis for modelling parenchymal patterns. In this chapter, we conduct experiments to evaluate the validity of all these methods for mammographic risk assessment. The full MIAS database (Suckling *et al.*, 1994) and a large subset of the DDSM database (Heath *et al.*, 2000) are used in the experiments. Mammographic images are classified into four categories according to the BIRADS density classification.

6.1 Experimental Data

There are a number of public and well-known databases for computer-aided mammographic image analysis, such as the Mammographic Image Analysis Society (MIAS) database (Suckling *et al.*, 1994), the Digital Database of Screening Mammography (DDSM) database (Heath *et al.*, 2000), and the European Prospective



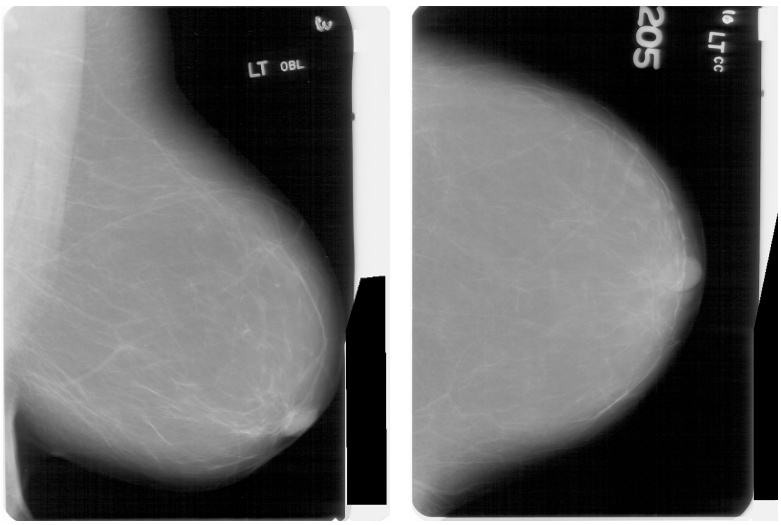
(a) (b)

Figure 6.1: Example mammographic images of a woman in the MIAS database: (a) right MLO mammogram ("mdb172rl"); (b) left MLO mammogram ("mdb171ll").

Investigation on Cancer (EPIC) database (Day *et al.*, 1999). To enable the comparison with existing work in the literature, we have used the MIAS database and the DDSM database in this chapter for evaluation (we did not use the EPIC database for evaluation here as we did not have ground truth for the BIRADS density classification).

6.1.1 MIAS Database

The MIAS database was generated by the Mammographic Image Analysis Society where mammograms were taken from the UK National Breast Screening Programme. The database contains 322 mammograms from 161 women. Bilateral (left and right) MLO view mammograms were taken for each woman. Mammograms were digitised by a Joyce-Loebl scanning microdensitometer which has a device linear in the optical density range 0-3.2 and represents each pixel with 8 bits. The spatial resolution of mammographic images is $50\mu m \times 50\mu m$ per pixel. The MIAS dataset provides detailed information about abnormalities (if present in mammograms), including the type of abnormalities (such as calcifications, well-defined/circumscribed masses, spiculated masses, ill-defined masses, architectural distortion, etc.), the severity of abnormalities (benign or malignant), and the location of abnormalities. Mammograms are categorised into three classes (i.e. fatty, fatty-glandular, and dense-glandular) according to the characteristics of background tissue. There is no breast density rating information available based on the BIRADS density classification. Figure 6.1 shows two example mammographic images ("mdb171ll" and "mdb172rl") of a woman in the MIAS database.



(a) (b)

Figure 6.2: Example mammographic images of a woman in the DDSM database: (a) right MLO mammogram ("D_4529_1.RIGHT_MLO"); (b) right CC mammogram ("p_4529_1.RIGHT_CC").

Three expert radiologists were involved to classify the 321 available mammographic images (mdb295ll was excluded for historical reasons) in the database into four classes according to the BIRADS density classification. The consensus between the three individual classification decisions was considered as the final class label. We use the consensus classification as the ground truth in this thesis. The consensus was the majority vote when two or three radiologists agree on a single class. If the three radiologists classified a mammogram as three different classes, the median result was selected as the consensus result. The consensus classification results of the MIAS database are 87 (27%) for BIRADS I, 103 (32%) for BIRADS II, 94 (29%) for BIRADS III, and 37 (12%) for BIRADS IV, respectively.

6.1.2 DDSM Database

The DDSM database was collected by the University of South Florida in collaboration with the Massachusetts General Hospital and Sandia National Laboratories. The database contains mammographic images for approximately 2500 women. For each woman, four mammograms were taken including both MLO and CC views of each breast. Mammograms were digitised by four scanners: DBA M2100 ImageClear (42 microns per pixel, 16 bits), Howtek 960 (43.5 microns per pixel, 12 bits), Lumisys 200 Laser (50 microns per pixel, 12 bits), and Howtek MultiRad850 (43.5 microns per pixel, 12 bits). Associated information is provided for each mammogram including patient information, breast density rating, and subtlety rating for abnormalities. In contrast to the MIAS database, breast density

ratings in the DDSM database are determined according to the four BIRADS density categories. We use the provided BIRADS classification as the ground truth in this thesis. We randomly select a large set of mammograms from the DDSM database for experiments, which consists of right MLO and CC mammograms of 831 women. The number of cases belonging to each BIRADS class is 106 (13%), 336 (40%), 255 (31%), and 134 (16%) for BIRADS I to BIRADS IV, respectively. Figure 6.2 shows two example mammographic images ("D_4529_1.RIGHT_MLO" and "p_4529_1.RIGHT_CC") of a woman in the DDSM database.

6.2 Evaluation Methodology

6.2.1 Classification Algorithm

We use a k-Nearest Neighbours (kNN) based classifier which combines the classic kNN classification and a distance weighted approach as described in Algorithm 2. The classification is initially based on a simple majority vote and a test object is classified by assigning the class label of the majority among its k nearest neighbours, unless multiple classes indicate the same number of training samples among the k nearest neighbours (i.e. a tie occurs in the decision making) in which case the distance weighted approach is applied to resolve the tie. It was shown by our experiments that when combined with the distance weighted approach the kNN classifier achieved approximately $< 1\%$ improvement of classification accuracy on average. The use of the kNN based classifier in this thesis is for the purpose of making a relatively fair comparison with the most closely related publications where kNN was used for the BIRADS breast density classification (e.g. Oliver *et al.* (2008); He *et al.* (2011, 2012)). The use of advanced classifiers or classifier ensembles is regarded as future work.

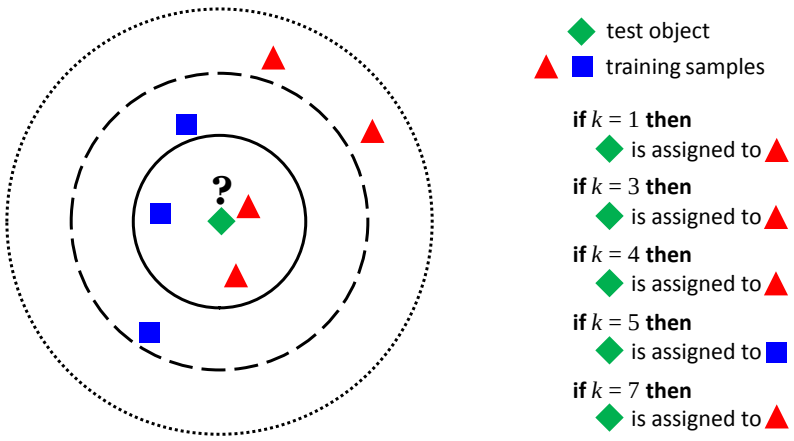


Figure 6.3: An example of the kNN classification.

Algorithm 2 The kNN based classifier combining the classic kNN and the distance weighted approach.

1. Compute the distances between the test object y and the training samples X to get the k nearest neighbours KNN:

$$KNN = \arg\min_{x_i \in X} \sum_{i=1}^{k} dist(x_i, y).$$

2. Get the number of training samples belonging to class C_i among the k nearest neighbours to y:

$$c_i(y) = |C_i \cap KNN|,$$

where $C_i \cap KNN$ denotes the set of neighbours belonging to class C_i among the k nearest neighbours, and $|\cdot|$ denotes the cardinality.

3. Determine the class label of y by the majority, i.e. the maximum of $c_i(y)$:

$$C(y) = \arg\max_i(c_i(y)).$$

4. If $C(y)$ indicates only one class has the most frequent label, then assign this label to y; else compute the sum of distances between y and the neighbours belonging to class C_i in $C(y)$, assign the label which has the minimal sum distance to y:

if $|C(y)| = 1$ **then**
  Output: $C(y)$
else
  $D(C_i) = \sum_{x_i \in C_i \cap KNN, C_i \in C(y)} dist(x_i, y)$
  $C(y) = \arg\min_i(D(C_i))$
  Output: $C(y)$
end if

where $D(C_i)$ is the sum of distances between y and the neighbours belonging to class C_i among the k nearest neighbours, and class C_i is one of the classes which have the same number of training samples among the k nearest neighbours.

To compute the distance between the test object y and the training sample x, we use two distance measures which are the Euclidean and Chi-square distance.

The Euclidean distance between x and y is calculated by:

$$\|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \tag{6.1}$$

where $\|\cdot\|$ is the Euclidean norm, which is the length of the vector connecting the two points x and y in the n-dimensional Euclidean space.

The Chi-square distance between x and y is calculated by:

$$\chi^2(x, y) = \frac{1}{2} \sum_{i=1}^{n} (x_i - y_i)^2/(x_i + y_i), \tag{6.2}$$

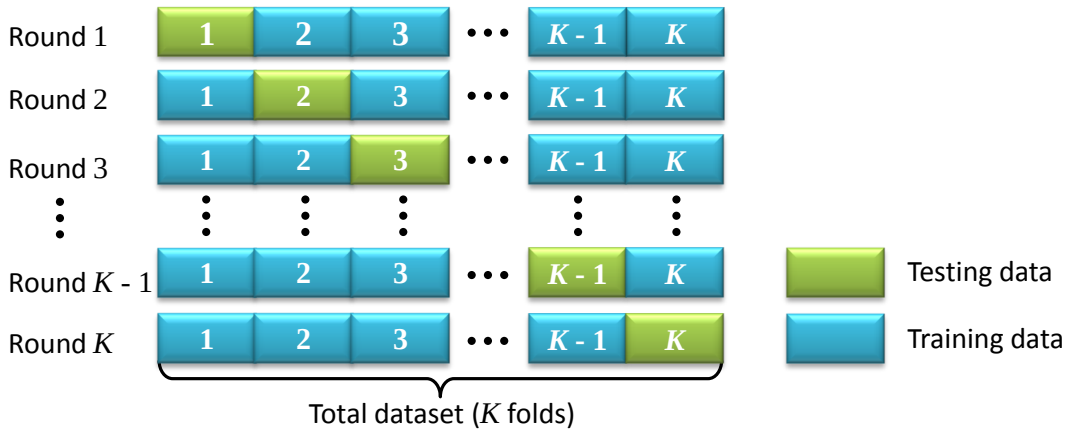


Figure 6.4: The diagram of the K-fold cross-validation.

where x and y often represent two frequency histograms, x_i and y_i are the frequencies at the i^{th} histogram bin.
An example of the kNN classification using Algorithm 2 is shown in Figure 6.3. If $k = 4$, the number of training samples belonging to each class is equal, and therefore the Euclidean distances between the test object and the four nearest neighbours are further considered to break the tie.

6.2.2 Cross-Validation Scheme

Cross-validation is a technique for estimating the performance and accuracy of a predictive model in practice (Geisser, 1993). For one round of cross-validation, the dataset is randomly partitioned into two complementary subsets. The predictive model is built based on one subset (i.e. the training set), and the validity of this model is tested using the other subset (i.e. the testing set). Multiple rounds of cross-validation are performed to reduce variability induced by the partition, and the validation results are averaged over the rounds.

6.2.2.1 K-Fold Cross-Validation

In K-fold cross-validation, the original dataset is randomly partitioned into K subsets. A single subset of the K subsets is used as the testing data for validating the model, and the remaining $K - 1$ subsets are used as the training data to generate the model. The cross-validation process is repeated K rounds so that each of the K subsets is used exactly once as the testing data. The K validation results corresponding to the K folds are then averaged (or otherwise combined) to produce a single estimation result. The diagram of the K-fold cross-validation is shown in Figure 6.4.

Table 6.1: A confusion matrix for the four-class classification.

		Automatic				
	Class	I	II	III	IV	Accuracy
Truth	I	n_{11}	n_{12}	n_{13}	n_{14}	n_{11}/n_1
	II	n_{21}	n_{22}	n_{23}	n_{24}	n_{22}/n_2
	III	n_{31}	n_{32}	n_{33}	n_{34}	n_{33}/n_3
	IV	n_{41}	n_{42}	n_{43}	n_{44}	n_{44}/n_4

6.2.2.2 Leave-One-Out Cross-Validation

In leave-one-out cross-validation, as its name suggests, a single sample of the original dataset is used as the testing data for validating the model, and all the remaining samples are used as the training data to generate the model. This process is repeated so that each sample in the dataset is used exactly once as the testing data. As described above, the leave-one-out cross-validation is the same as the K-fold cross-validation with a K equal to the number of samples in the original dataset. In this thesis, we mainly use a leave-one-woman-out cross-validation methodology. When classifying one mammogram, the other mammogram from the same woman is excluded from the training data to avoid bias (left and right mammograms from a single woman tend to have similar tissue features).

6.2.3 Result Representation

As mentioned above, the consensus BIRADS density rating from the three expert radiologists is used as the ground truth for the MIAS database, and the BIRADS density rating information associated with the DDSM database is used as the ground truth for the DDSM database. The validity of the proposed methods for mammographic risk assessment is evaluated by comparing the automatic classification with the ground truth. Classification results are represented in the form of confusion matrices. Table 6.1 shows an example confusion matrix for the four-class classification. Each row of the matrix represents the samples in an actual class according to the ground truth, while each column of the matrix represents the samples in a predicted class determined by the automatic classification. In Table 6.1, n_{ij} $(i, j = 1, 2, 3, 4)$ denotes the number of samples belonging to class i and classified into class j. When $i = j$, n_{ij} corresponds to the number of correctly classified samples, while when $i \neq j$, n_{ij} corresponds to the number of incorrectly classified samples. $n_i = \sum_{j=1}^{4} n_{ij}$ $(i = 1, 2, 3, 4)$ denotes the number of samples in each actual class. Thus, the classification accuracy for each class is n_{11}/n_1, n_{22}/n_2, n_{33}/n_3, and n_{44}/n_4, respectively. The overall classification accuracy for the four-class classification is calculated by $\sum_{i=1}^{4} n_{ii} / \sum_{i=1}^{4} n_i$.

6.3 Evaluating the Proposed Methods

In the above sections, the detailed information on the data used for experiments and the methodologies used for evaluation has been presented. In this section, we evaluate the methods proposed previously in terms of the validity for mammographic risk assessment.

6.3.1 Evaluation of Breast Density Segmentation

Two breast density segmentation methods were proposed in Chapter 3, which were based on a modified fuzzy c-means (MFCM) algorithm and a topographic representation, respectively. Here, we investigate the capability of the two methods in breast density classification, which is performed based on the features generated from the segmentation results.

6.3.1.1 Evaluation of MFCM Based Segmentation

The evaluation of the MFCM based breast density segmentation method was based on the full MIAS database. For each mammogram, the breast region was first segmented using the approach proposed in Section 3.2. Following breast region segmentation, the resulting breast region was filtered with the adaptive local window filter defined in Section 3.3.2, in order to eliminate the sensitivity to noise and small intensity inhomogeneity effects. The same settings of the parameters in the algorithm were used: 5×5 local windows, $\lambda_s = 1$, and $\lambda_g = 3$. After that, 40 mammographic images were selected as a training set for analysing the general distribution of breast tissue density in the database. The 40 training images were selected in two ways: one (denoted by Train I) is choosing 10 images from each BIRADS class; the other (denoted by Train II) is randomly sampling 40 images from all the 321 images in the database regardless of the BIRADS class label. A greylevel histogram was generated from all the training images, describing the overall intensity distribution of the pixels in the training set. The MFCM based clustering was performed based on this histogram to partition the pixels in the training set into N_d clusters. As such, each greylevel had N_d membership values with respect to the N_d clusters. The breast region can be segmented into N_d sub-regions by assigning each pixel within the breast region into one of the N_d clusters according to the maximum membership value of its greylevel. Thus, the N_d sub-regions of the breast have different tissue densities since they correspond to different greylevel sections of the full greylevel range. The process of the MFCM based breast density segmentation and some representative segmentation results (the number of sub-regions $N_d = 4$) are shown in Figure 6.5. The colour of the

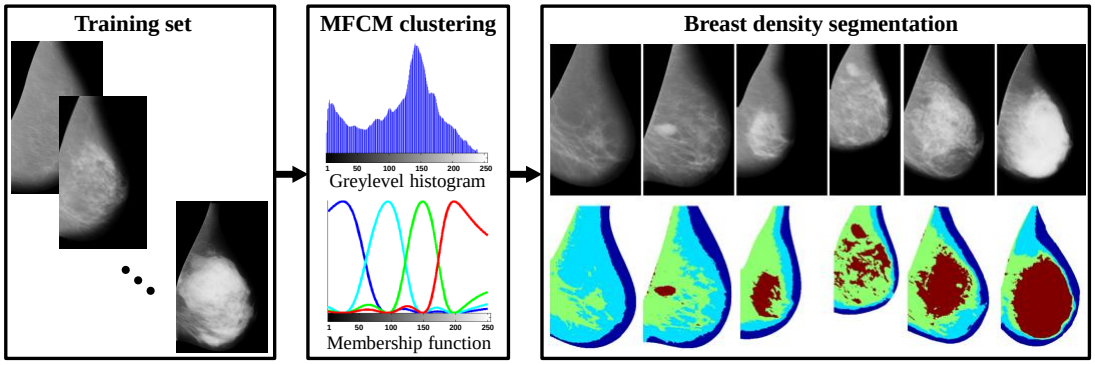


Figure 6.5: The process of the MFCM based breast density segmentation.

sub-regions in the breast is consistent with the colour of the corresponding cluster according to the maximum membership value. The density of the four sub-regions increases in the order of blue, light blue, green, and red.

The relative areas of the resulting sub-regions with respect to the area of the whole breast region were used as the density based features for breast density classification. The kNN based classifier and the leave-one-woman-out cross-validation methodology were used for the classification. The χ^2 distance was used to measure the dissimilarity between two feature vectors. The overall classification accuracy (CA) was used as the evaluation metric. For each selection way (Train I/II), 10 training sets were created for investigating the variability in the classification accuracy due to different training images. The classification performance was also tested with respect to the number of the sub-regions (N_d).

A range of values for N_d from 2 to 10 were tested, and thus the corresponding feature space dimensionality varied from 2 to 10. Table 6.2 shows the obtained classification accuracies when using different values of N_d and different training sets. Figure 6.6 shows the means and standard deviations of classification accuracy over the ten training sets for Train I and Train II with respect to different values of N_d. As can be seen from this figure, the classification accuracy for the two groups of training sets has a considerable increase as the number of the sub-regions increases from 2 to 5 and tends to be stable when $N_d \geq 5$ with a very slight decrease for $N_d = 7$. The best CAs for Train I and Train II were obtained when $N_d = 6$, which were $66.95 \pm 0.63\%$ and $67.60 \pm 1.25\%$, respectively. The two groups of training sets provided similar results for the same value of N_d. Table 6.3 shows two confusion matrices for two training sets respectively from Train I and Train II when $N_d = 6$.

Table 6.2: Classification accuracies for different values of N_d and training sets.

	Train I (10 for each class)									
N_d	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
2	28.35	24.30	28.35	26.79	26.79	27.73	27.73	28.97	28.04	28.04
3	53.58	49.84	60.12	51.09	56.07	56.39	51.40	52.02	50.47	52.02
4	63.24	60.75	61.99	60.75	61.68	62.93	63.55	61.68	61.37	61.99
5	66.98	66.04	66.98	64.80	67.91	67.29	65.73	67.60	66.98	66.98
6	66.98	68.22	66.36	66.67	66.67	66.04	67.60	67.29	66.98	66.67
7	64.17	64.17	65.73	64.49	64.80	63.55	65.42	65.42	64.49	63.86
8	66.36	65.11	66.04	66.36	68.22	66.98	66.67	66.36	66.36	66.36
9	65.73	66.67	66.04	67.60	68.54	65.42	67.91	65.73	66.04	65.73
10	66.04	66.04	64.80	66.04	66.36	66.04	66.98	66.67	65.42	66.04

	Train II (random sampling)									
N_d	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
2	27.73	28.97	24.92	26.79	27.73	26.79	24.30	28.35	26.79	26.79
3	53.58	47.35	48.91	51.09	47.35	47.04	48.60	54.21	52.65	45.17
4	62.62	61.68	60.44	61.99	59.81	61.37	60.44	61.37	61.99	58.57
5	67.60	66.04	65.73	66.04	65.73	64.80	66.36	66.67	66.67	62.62
6	66.04	66.67	67.91	68.85	68.54	68.85	69.16	65.73	66.67	67.60
7	64.49	66.36	65.42	67.60	65.11	65.11	67.29	65.42	64.49	68.22
8	65.42	65.73	65.11	65.42	65.73	65.42	66.98	67.60	66.04	66.98
9	65.11	67.60	67.29	66.04	66.36	66.67	67.60	66.98	66.67	67.29
10	66.36	66.36	66.04	66.36	66.98	66.04	66.67	65.11	66.04	67.29

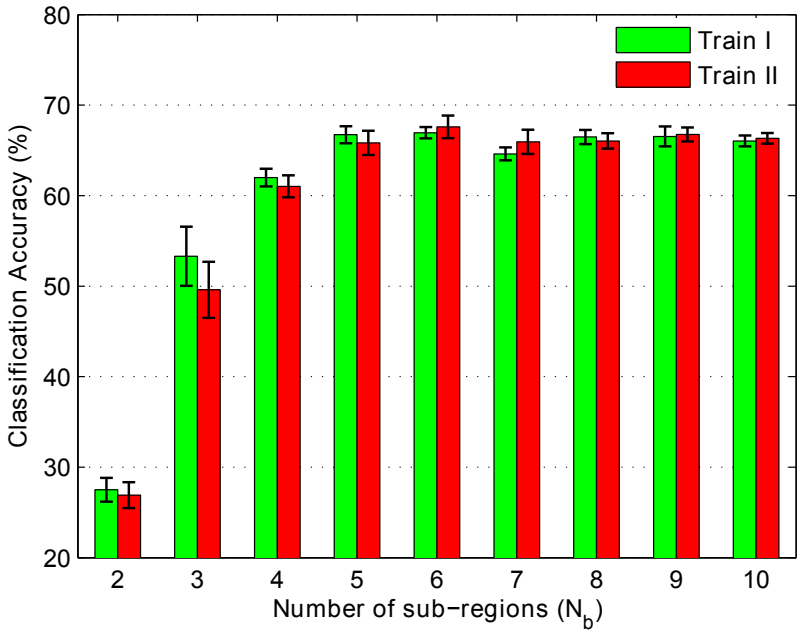


Figure 6.6: The means and standard deviations of classification accuracy over the ten training sets for Train I and Train II.

Table 6.3: Two confusion matrices for two training sets ($N_d = 6$), with the overall classification accuracy equal to 68.22% and 69.16%, respectively.

(a) Set 2 of Train I (CA = 68.22%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	73	13	1	0	84%
	II	11	71	20	1	69%
	III	0	32	56	6	60%
	IV	0	3	15	19	51%

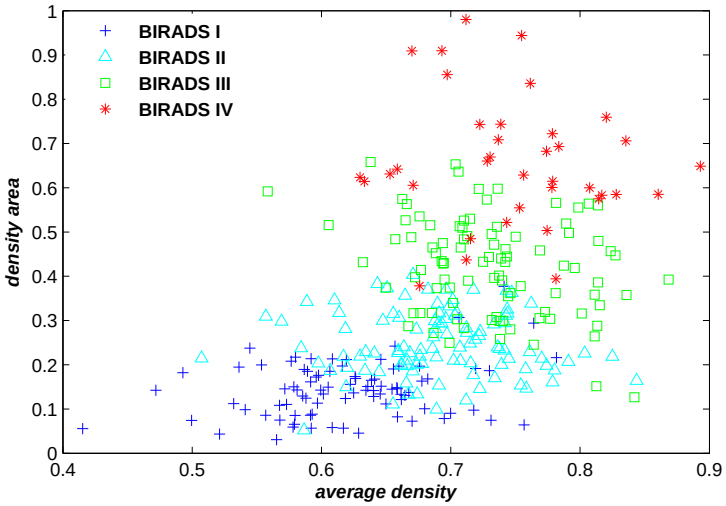
(b) Set 7 of Train II (CA = 69.16%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	72	13	2	0	83%
	II	14	70	18	1	68%
	III	0	23	62	9	66%
	IV	0	3	16	18	49%

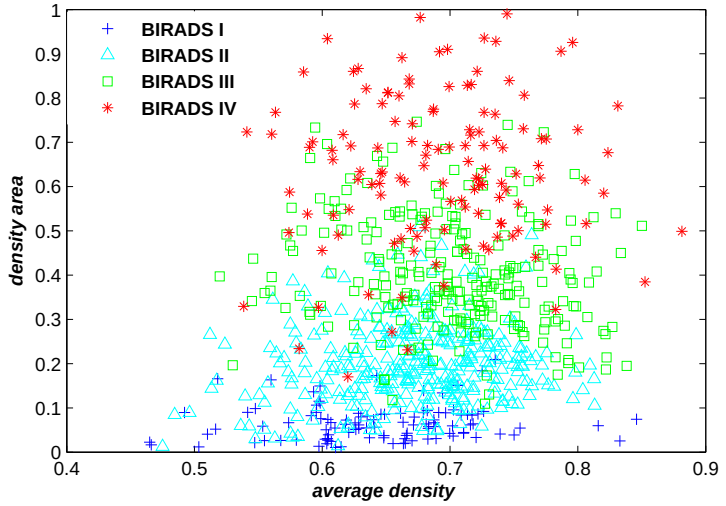
6.3.1.2 Evaluation of Topography Based Segmentation

The full MIAS database (containing 321 MLO mammograms) and the large subset of the DDSM database (containing 831 MLO and CC mammograms) were used for evaluating the topography based breast density segmentation method. As described in Section 3.4, the breast region was first segmented from each mammogram. Eighty-five intensity levels were then used to create a topographic map of the breast region. Subsequently, dense tissue regions were segmented from the breast by detecting prominent/independent *shapes* from the topographic map. Finally, a density map was generated based on the resulting dense tissue regions. Two features were derived from the density map for breast density classification: one is *density area* representing the total area of the segmented dense tissue regions; the other is *average density* calculated by averaging the intensity values in the density map, representing the average density of the segmented dense tissue regions. Scatter plots of the two features for the two databases are shown in Figure 6.7. The value of *density area* was normalised by dividing the area of the breast region to avoid bias induced by the size of individual breasts and as such the normalised *density area* represented the relative proportion of dense tissue in the breast. On the other hand, the value of *average density* was normalised by dividing the maximum intensity value of 255 within the full greylevel scale. The purpose of using a uniform value to normalise *average density* instead of a varied value according to mammograms was to preserve the original correlation of intensity information between mammograms.

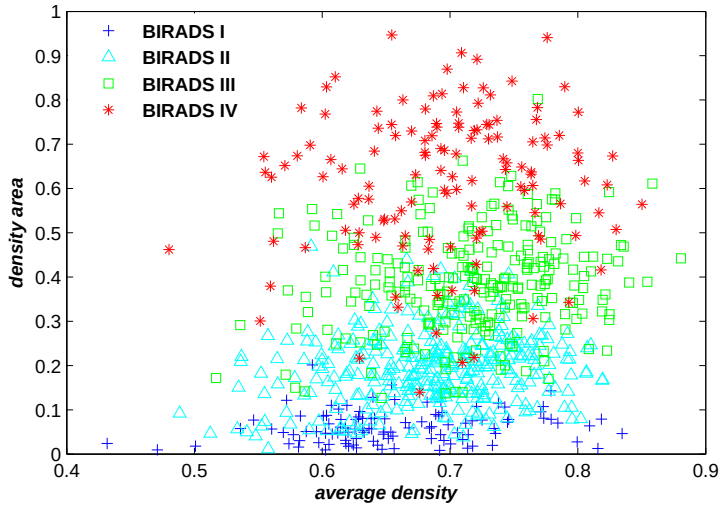
The kNN based classifier described earlier was employed for the classification of mammograms. The Euclidean distance was used as the similarity measure. The leave-one-woman-out methodology was used for cross-validation. For the MIAS database, when classifying one MLO mammogram of one woman, the opposite-side mammogram was excluded from the training samples. For the DDSM database, we classified mammograms of the MLO view and the CC view separately. Thus, the leave-one-woman-out methodology was regarded as the leave-one-image-out



(a) 321 MLO mammograms in MIAS



(b) 831 MLO mammograms in DDSM



(c) 831 CC mammograms in DDSM

Figure 6.7: Scatter plots of the two features for mammograms in the two databases.

Table 6.4: Classification accuracies for the MIAS and DDSM databases when using each or both of the derived features, where *feature* 1 and *feature* 2 represent *density area* and *average density*, respectively.

Feature	MIAS	DDSM (MLO)	DDSM (CC)
feature 1	71.03%	74.37%	72.68%
feature 2	50.47%	44.89%	41.88%
feature 1/2	76.01%	75.21%	74.61%

cross-validation for mammograms of each view.

We tested the individual performance of the two features and the performance of the combination of the two features for breast density classification. Table 6.4 shows the classification accuracies of the four-class BIRADS classification for the MIAS and DDSM databases when using each or both of the two features. It is shown that using the combination of the two features produced the best performance compared with using any individual feature, and *density area* played a main role in the classification. The poor performance of *average density* in particular for the DDSM database could be resulted from the fact that there is a variability in the intensity of breast tissue having the same density among different mammograms and the DDSM database was generated using different digitisers which possibly induced a larger variability.

The resulting confusion matrices when using both features for the BIRADS breast density classification can be found in Table 6.5, where Table 6.5(a) shows the confusion matrix for the full MIAS database with the overall classification accuracy of 76.01%, Tables 6.5(b) and 6.5(c) show the confusion matrices for the DDSM database with the overall classification accuracy equal to 75.21% and 74.61% for the MLO view and the CC view, respectively. The overall classification accuracy for the DDSM database of the individual view was slightly worse than that for the MIAS database. This might be due to the fact that the DDSM database used in the experiments contained more mammograms than the MIAS database and the mammograms belonging to the same class showed higher variance than the MIAS database. However, the obtained results for the DDSM database still indicate the robustness of our method on a large dataset. When considering each BIRADS class, the performance for a single BIRADS class was different between the two databases. The best performance was achieved by BIRADS I for the MIAS database while BIRADS II indicated the best results for the DDSM database. This might be due to the different distribution of the four BIRADS classes in the two databases, which resulted in a different distribution over the four classes in the training samples when using the leave-one-out cross-validation methodology. When focusing on the DDSM database, the MLO and CC views also indicated

Table 6.5: Confusion matrices for the MIAS and DDSM databases using the two features of *density area* and *average density* and leave-one-woman-out.

(a) MIAS (CA = 76.01%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	72	13	2	0	83%
	II	18	73	12	0	71%
	III	0	19	70	5	74%
	IV	0	0	8	29	78%

(b) MLO DDSM (CA = 75.21%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	80	26	0	0	75%
	II	33	266	35	2	79%
	III	0	40	186	29	73%
	IV	0	4	37	93	69%

(c) CC DDSM (CA = 74.61%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	72	34	0	0	68%
	II	29	280	27	0	83%
	III	0	54	180	21	71%
	IV	0	5	41	88	66%

(d) MLO/CC DDSM (CA = 81.23%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	93	13	0	0	88%
	II	27	286	23	0	85%
	III	0	41	199	15	78%
	IV	0	5	32	97	72%

diverse results over the four BIRADS classes. For BIRADS I, better performance was obtained by using the MLO view; for BIRADS II, the CC view outperformed the MLO view; while for BIRADS III and IV, the MLO view provided better results than the CC view. Therefore, we investigated the capability of combining the MLO and CC views of each woman for breast density classification. The two features extracted from each individual view were concatenated into a single feature vector for each woman. The combination of the two views provided a good improvement as shown in Table 6.5(d). An overall classification accuracy of 81.23% was obtained and the classification accuracies for the four BIRADS classes were all improved, which indicated a better performance than using the individual views. This confirmed the fact that the MLO view mammogram and the CC view mammogram of the same woman can provide complementary information for mammographic risk assessment.

The feature of *average density* was calculated by averaging the pixel values in the density map. As described in Section 3.4.3, the pixels belonging to the same *shape* in the density map have a single intensity value which is the average intensity of the *shape*. Thus, the greylevel resolution of the density map is reduced such that each dense tissue region is represented as a uniform region. We also investigated the performance of using the original pixel values derived from the original mammographic images to calculate *average density*. Slightly worse results were produced by using the same *density area* and the recalculated *average density*. The resulting classification accuracies were 72.90%, 73.41%, 72.56%, and 80.51% for MIAS, DDSM (MLO), DDSM (CC), and DDSM (MLO/CC), respectively.

Table 6.6: The average confusion matrices for the DDSM database using the two features of *density area* and *average density* and 10-fold cross-validation.

(a) MLO DDSM (CA = $75.09 \pm 0.62\%$)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	78.9±1.4	27.1±1.4	0.0±0.0	0.0±0.0	74.4±1.3%
	II	32.3±1.6	266.4±2.8	35.8±2.3	1.5±0.7	79.3±0.8%
	III	0.0±0.0	40.1±1.4	185.8±1.5	29.1±1.6	72.9±0.6%
	IV	0.0±0.0	4.3±0.5	36.8±2.6	92.9±2.4	69.3±1.8%

(b) CC DDSM (CA = $74.34 \pm 0.72\%$)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	72.8±1.9	33.2±1.9	0.0±0.0	0.0±0.0	68.7±1.8%
	II	30.8±2.4	275.7±3.5	29.5±2.0	0.0±0.0	82.1±1.1%
	III	0.0±0.0	54.1±0.9	179.8±1.7	21.1±1.6	70.5±0.7%
	IV	0.0±0.0	5.8±0.6	38.7±1.8	89.5±1.7	66.8±1.3%

(c) MLO/CC DDSM (CA = $80.88 \pm 0.44\%$)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	91.5±1.4	14.5±1.4	0.0±0.0	0.0±0.0	86.3±1.3%
	II	27.0±1.5	285.1±2.4	23.9±1.8	0.0±0.0	84.9±0.7%
	III	0.0±0.0	40.5±1.6	199.2±2.9	15.3±1.9	78.1±1.1%
	IV	0.0±0.0	5.1±0.3	32.6±1.2	96.3±1.3	71.9±0.9%

In addition, for the DDSM database, we investigated the classification performance of *density area* and *average density* using the stratified 10-fold cross-validation (10-FCV) methodology. To avoid bias induced by the random partition, ten rounds of 10-FCV were performed. The average classification accuracies were $75.09 \pm 0.62\%$, $74.34 \pm 0.72\%$, and $80.88 \pm 0.44\%$ for the MLO view, the CC view, and the combination of the MLO and CC views, respectively. Note that the standard deviations above were computed over the ten rounds and the standard deviations over the 100 classifier models (10 folds $\times$ 10 rounds) were 3.98%, 4.72% and 4.12%. When comparing with the classification accuracies achieved by leave-one-woman-out, equivalent results were indicated by 10-FCV. The average confusion matrices over the ten rounds of 10-FCV are shown in Table 6.6.

6.3.2 Evaluation of Breast Tissue Appearance Modelling

A scheme for mammographic tissue appearance modelling based on local features was proposed in Chapter 4, where five strategies were investigated by employing different types of local features. Mammographic tissue was modelled based on

Table 6.7: Classification accuracy for the five strategies/different parameters.

Approach	Overall classification accuracy for different parameters		
LBP	59.19 (3×3 neighbours)	**60.12** (5×5 neighbours)	47.04 (7×7 neighbours)
LGA	67.91 (3×3, $N_g = 8$)	70.72 (3×3, $N_g = 16$)	**72.27** (5×5, $N_g = 16$)
BIF	**70.40** ($\sigma_{base} = 1$, $\varepsilon = 0$)	66.04 ($\sigma_{base} = 1.2$, $\varepsilon = 0$)	67.91 ($\sigma_{base} = 1$, $\varepsilon = 200$)
Texton I	66.98 (40 textons)	68.54 (80 textons)	**69.78** (160 textons)
Texton II	66.98 (40 textons)	69.78 (80 textons)	**74.77** (160 textons)

statistical analysis of local tissue appearance. As such, mammographic images can be represented as occurrence histograms over a generic dictionary containing a summary of local features. In this section, we apply the resulting histogram models to breast density classification. The performance is evaluated using the full MIAS database based on the BIRADS density classification.

The occurrence histograms were normalised using L_1 normalisation. The leave-one-woman-out evaluation methodology was used for the classification. When classifying the left MLO mammogram, the right MLO mammogram from the same woman was excluded from the training samples. The kNN based classifier described above was used for the classification which was initially based on a simple majority vote, unless a tie occurred when making the decision, in which case a distance weighted approach was involved. The dissimilarity between two histogram models was measured using the χ^2 distribution comparison.

As described in Sections 4.2 and 4.3, there are a number of parameters in the five modelling strategies which may have important effects on the resulting models. This has been indicated by the resulting label maps by using different parameters (see examples in Section 4.3). Here, we tested the performance of each individual strategy with respect to different parameters. Table 6.7 shows the overall classification accuracies for the five modelling strategies with respect to different parameters. LBP obtained the best classification accuracy of 60.12% when using 5×5 neighbours. For LGA, the best classification accuracy was 72.27% with 5×5 neighbours and a greylevel resolution of 16. The best classification accuracy for BIF was 70.40% when setting $\sigma_{base} = 1$ and $\varepsilon = 0$. Different values of ε produced similar classification results. For Texton I and Texton II, the best classification accuracy was obtained when generating 160 textons, which was 69.78% and 74.77%, respectively. As indicated above, the best classification performance was achieved by Texon II, which was followed by LGA, providing the second-best result. BIF and Texton I, which are both filter bank based approaches, produced similar results. LBP performed worst among these five approaches especially when using a large size of local windows.

The different performance of the five strategies can be explained by their individual properties. LBP transforms local tissue appearance into an *LBP* number,

Table 6.8: Confusion matrices for breast density classification when using the individual strategies for breast tissue appearance modelling ((a) - (e)) and using the combination of the five strategies ((f)).

(a) LBP (CA = 60.12%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	67	19	1	0	77%
	II	26	39	38	0	38%
	III	1	23	65	5	69%
	IV	0	0	15	22	59%

(b) LGA (CA = 72.27%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	78	8	1	0	90%
	II	22	73	8	0	71%
	III	0	30	62	2	66%
	IV	1	0	17	19	51%

(c) BIF (CA = 70.40%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	76	9	2	0	87%
	II	29	57	17	0	55%
	III	2	23	66	3	70%
	IV	0	1	9	27	73%

(d) Texton I (CA = 69.78%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	78	8	1	0	90%
	II	18	68	17	0	66%
	III	0	29	63	2	67%
	IV	0	4	18	15	41%

(e) Texton II (CA = 74.77%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	76	10	1	0	87%
	II	9	81	13	0	79%
	III	0	24	66	4	70%
	IV	1	0	19	17	46%

(f) Combination (CA = 77.88%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	82	5	0	0	94%
	II	18	75	10	0	73%
	III	0	21	71	2	76%
	IV	0	0	15	22	59%

excluding greylevel information from the process of breast tissue modelling. However, greylevel information plays an important role in describing breast tissue density in mammographic images. Therefore, this resulted in its worst behavior among the five approaches. In contrast, LGA and Texton II focus on greylevel appearance within local neighbourhoods. As shown in Figures 4.8, 4.10 and 4.18, the resulting label maps indicate density based segmentation. This might explain their superior performance for breast density classification. On the other hand, BIF and Texton I are based on filter banks and include greylevel information into the modelling process. However, they are sensitive to small tissue structures, such as small edges, bars, and blobs, which are relatively less significant for evaluating the overall breast density, thus they indicated the intermediate results.

The confusion matrices corresponding to the best classification accuracies for the five strategies (marked in bold in Table 6.7) are shown in Table 6.8. For BIRADS I, the best classification accuracy was 90% obtained by LGA and Texton I. For BIRADS II, Texton II provided the best classification accuracy of 79%. For BIRADS III, BIF and Texton II indicated the best classification accuracy of 70%. Finally, for BIRADS IV, BIF achieved the best classification accuracy of 73%.

In addition, we investigated the correlation between every two groups of classification results obtained by two of the five strategies. The comparison results are also shown in the form of confusion matrices, where the number in the i^{th} row and j^{th} column represents the number of mammograms classified into class i and class j by the two approaches, respectively. Table 6.9 shows the ten resulting confusion matrices after comparing every two of the five strategies. A strong correlation was indicated between LGA and Texton II with an overall agreement of 88%, as they both mainly focused on local intensity information. This was confirmed by their resulting label maps which indicated similar segmentation results. In addition, LGA and Texton I, and Texton I and Texton II showed a relatively strong correlation with an overall agreement of 75% and 73%, respectively. On the other hand, the correlations between all the other pairs of strategies were relatively low and the overall agreement varied from 58% to 68%. This showed the diversity of the classification results among the five strategies. Thus, we further investigated the performance of combining the outputs of the five individual classifiers.

The binary classification results of the kNN classifiers were transformed into continuous-valued outputs, which can be interpreted as posterior probabilities. For a test object y, the probability of it belonging to class C_i is proportional to the number of neighbours belonging to class C_i among the k nearest neighbours, and the probability value has an inverse relationship with the distances between y and these neighbours, which is computed by $P(C_i \mid y) = \sum_{x_i \in C_i \cap KNN} \frac{1}{1+dist(x_i,y)}$, where $C_i \cap KNN$ denotes the neighbours belonging to class C_i among the k nearest neighbours. If $C_i \cap KNN = \phi$, $P(C_i \mid y) = 0$. The values of $P(C_i \mid y)$ $(i = 1, 2, 3, 4)$ are normalised to make the sum of the outputs over all the four BIRADS classes equal to 1.

A weighted average combination rule was used to compute the total probability for each class. For a test mammogram, the total probability corresponding to class C_i (denoted by $P_{sum}(C_i)$) is obtained by $P_{sum}(C_i) = \sum_{t=1}^{5} w_t P(C_i \mid y_t)$, where $P(C_i \mid y_t)$ is the output of the t^{th} classifier, and w_t is the corresponding weight value. We set $w_1 = 0.6$, $w_2 = 0.4$, $w_3 = 0.7$, $w_4 = 0.1$, and $w_5 = 0.9$ experimentally, but small variations provided similar results. The obtained overall classification accuracy was 77.88% for the four BIRADS density categories, which indicated better performance compared with those obtained by the individual classifiers. The confusion matrix is shown in Table 6.8(f). Note that none of mammograms were mis-classified by more than one BIRADS class. For two-class (low/high) density classification, the classification accuracy increased to 90.34%.

Table 6.9: Confusion matrices for the comparison of the five groups of classification results of the five different strategies. AG denotes the overall agreement between every two strategies.

(a) LGA vs LBP (AG = 61%)

		LGA				
	BIRADS	I	II	III	IV	AG
LBP	I	77	15	2	0	82%
	II	20	47	14	0	58%
	III	3	49	59	8	50%
	IV	1	0	13	13	48%

(b) BIF vs LBP (AG = 59%)

		BIF				
	BIRADS	I	II	III	IV	AG
LBP	I	76	16	2	0	81%
	II	26	35	20	0	43%
	III	5	39	62	13	52%
	IV	0	0	10	17	63%

(c) Texton I vs LBP (AG = 62%)

		Texton I				
	BIRADS	I	II	III	IV	AG
LBP	I	71	22	1	0	76%
	II	22	45	13	1	56%
	III	3	41	71	4	60%
	IV	0	1	14	12	44%

(d) Texton II vs LBP (AG = 58%)

		Texton II				
	BIRADS	I	II	III	IV	AG
LBP	I	68	23	3	0	72%
	II	16	46	19	0	57%
	III	1	46	62	10	52%
	IV	1	0	15	11	40%

(e) BIF vs LGA (AG = 68%)

		BIF				
	BIRADS	I	II	III	IV	AG
LGA	I	84	14	2	1	83%
	II	21	60	30	0	54%
	III	2	16	57	13	65%
	IV	0	0	5	16	76%

(f) Texton I vs LGA (AG = 75%)

		Texton I				
	BIRADS	I	II	III	IV	AG
LGA	I	86	11	3	1	85%
	II	9	78	24	0	70%
	III	1	18	65	4	74%
	IV	0	2	7	12	57%

(g) Texton II vs LGA (AG = 88%)

		Texton II				
	BIRADS	I	II	III	IV	AG
LGA	I	85	16	0	0	84%
	II	1	96	14	0	86%
	III	0	3	83	2	94%
	IV	0	0	2	19	90%

(h) Texton I vs BIF (AG = 65%)

		Texton I				
	BIRADS	I	II	III	IV	AG
BIF	I	80	22	5	0	75%
	II	14	55	20	1	61%
	III	2	28	61	3	65%
	IV	0	4	13	13	43%

(i) Texton II vs BIF (AG = 65%)

		Texton II				
	BIRADS	I	II	III	IV	AG
BIF	I	74	30	3	0	69%
	II	9	60	21	0	67%
	III	2	25	60	7	64%
	IV	1	0	15	14	47%

(j) Texton II vs Texton I (AG = 73%)

		Texton II				
	BIRADS	I	II	III	IV	AG
Texton I	I	79	16	1	0	82%
	II	5	75	27	2	69%
	III	1	24	67	7	68%
	IV	1	0	4	12	71%

6.3.3 A Combined Modelling of Breast Tissue

In the previous two sections, we have evaluated the validity of the two breast density segmentation methods and the five breast tissue appearance modelling strategies for breast density classification. In these approaches, breast density segmentation focused on breast tissue density while breast tissue appearance modelling focused on breast tissue patterns. Here, we propose to model breast tissue covering both density and tissue patterns. For breast tissue density modelling, we use the MFCM based approach where the overall distribution of breast tissue density is analysed based on a greylevel histogram and the breast region can be segmented into a number of density sub-regions according to the resulting membership functions. For breast tissue pattern modelling, we use the texton based approach (Texton II) where tissue patterns are modelled using a set of textons generated from local image patches and the global appearance of breast tissue can be represented as an occurrence histogram over the textons.

To generate a combined model of breast tissue incorporating both breast tissue density and patterns, we concatenate the relative proportions of the density sub-regions and the texton histogram of the breast into a single feature vector, and therefore the dimensionality of the feature space is determined by how many sub-regions the breast is segmented and how many textons are used to model breast tissue appearance. Thus, breast tissue is represented with the joint distribution of breast tissue density and local breast tissue appearance.

The full MIAS database was used to test the capability of the combined models of breast tissue for mammographic risk assessment. We segmented the breast region into 8 sub-regions and used 160 3×3 textons to identify local tissue appearance. Thus, the resulting breast tissue model consisted of 168 features. We employed the kNN based classifier for the classification and the χ^2 distance was used to measure the dissimilarity between the breast tissue models. All features in the models were normalised to the range of $[0, 1]$. A stratified two-fold cross-validation methodology was used for evaluation. The two folds were stratified so that they contained approximately the same proportions of the four BIRADS classes as the original database. As a result, Fold I contained 161 mammograms and Fold II contained 160 mammograms due to the odd number of mammograms contained in the MIAS database. Five rounds of two-fold cross-validation were performed to avoid bias induced by the partition of the database. Table 6.10 shows the classification accuracies of the four-class BIRADS classification for density based features, tissue pattern based features, and the combined features based on both density and tissue patterns. As can be seen from the table, using the combined features covering both density and tissue patterns provided better performance than using the features

Table 6.10: Classification accuracies for density based features, tissue pattern based features, and the combined features based on both density and tissue patterns.

Feature	Round 1	Round 2	Round 3	Round 4	Round 5
density	63.24%	65.72%	65.43%	67.29%	65.74%
pattern	68.85%	68.85%	70.41%	68.85%	68.24%
density/pattern	69.16%	69.77%	70.72%	70.41%	69.79%

based solely on one aspect. The obtained classification accuracy was $69.97 \pm 0.61\%$. The tissue pattern based features derived from breast tissue appearance modelling produced better results than the density based features derived from breast density segmentation. Their resulting classification accuracies were $69.04 \pm 0.81\%$ and $65.48 \pm 1.45\%$, respectively (the standard deviation was calculated over the five rounds). This indicated slightly worse performance than that of the MFCM based segmentation when $N_b = 8$ and the Texton II approach when using 160 textons presented in the previous two sections. This might be explained by the fact that the two-fold cross-validation methodology was used here for evaluation where the classifier model was built based on a smaller training set than that of the leave-one-woman-out cross-validation methodology. The resulting confusion matrices of the density/pattern features are shown in Table 6.11 (left column). BIRADS I and II showed better performance than BIRADS III and IV. The best performance was achieved for BIRADS I, implying the resulting breast tissue models are the most discriminative to disdinguish mammograms in BIRADS I. By contrast, BIRADS IV showed the worst performance, having a big confusion in discriminating between BIRADS III and IV.

In the above classification process, the kNN classifier weighted all the features equally, without taking into account their discriminating capability. To select a set of discriminative features from the density/pattern based features and reduce the dimensionality of the constructed breast tissue models, we applied the sequential forward selection (SFS) algorithm (Kittler, 1986), which is a well-known feature selection technique. The SFS algorithm starts from an empty feature set and sequentially adds the feature which maximises the objective function when combined with the feature set already selected in the previous iterations, which has a tendency to get trapped at local minima and therefore seldom succeeds in reaching the global optimum. Here, we selected 40 features which were the 40 top-ranking features sequentially selected by SFS. In the feature selection process, SFS was performed solely based on the training set excluding the test object to avoid bias. As a result, ten feature sets were selected (two-fold cross-validation, five rounds). Figure 6.8 shows the number of times each feature was selected over

Table 6.11: Confusion matrices of the density/pattern features for the five rounds of the two-fold cross-validation. Left column: using the full feature space; right column: using the reduced feature space.

(a) Round 1 (CA = 69.16%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	73	13	1	0	84%
	II	13	70	20	0	68%
	III	0	32	61	1	65%
	IV	1	0	18	18	49%

(b) Round 1 (CA = 71.96%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	73	13	1	0	84%
	II	9	77	17	0	75%
	III	1	25	62	6	66%
	IV	1	1	16	19	51%

(c) Round 2 (CA = 69.77%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	70	15	2	0	80%
	II	10	77	16	0	75%
	III	0	28	61	5	65%
	IV	0	2	19	16	43%

(d) Round 2 (CA = 70.40%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	72	15	0	0	83%
	II	11	72	19	1	70%
	III	0	22	67	5	71%
	IV	1	1	20	15	41%

(e) Round 3 (CA = 70.72%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	67	19	1	0	77%
	II	12	81	10	0	79%
	III	0	26	64	4	68%
	IV	1	2	19	15	41%

(f) Round 3 (CA = 70.09%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	67	19	1	0	77%
	II	11	77	15	0	75%
	III	0	23	65	6	69%
	IV	1	1	19	16	43%

(g) Round 4 (CA = 70.41%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	75	12	0	0	86%
	II	10	75	18	0	73%
	III	1	29	58	6	62%
	IV	0	2	17	18	49%

(h) Round 4 (CA = 71.33%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	72	15	0	0	83%
	II	9	81	12	1	79%
	III	0	31	58	5	62%
	IV	0	3	16	18	49%

(i) Round 5 (CA = 69.79%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	72	14	1	0	83%
	II	8	75	20	0	73%
	III	0	31	58	5	62%
	IV	1	2	15	19	51%

(j) Round 5 (CA = 71.96%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	74	12	1	0	85%
	II	9	76	18	0	74%
	III	0	26	61	7	65%
	IV	0	2	15	20	54%

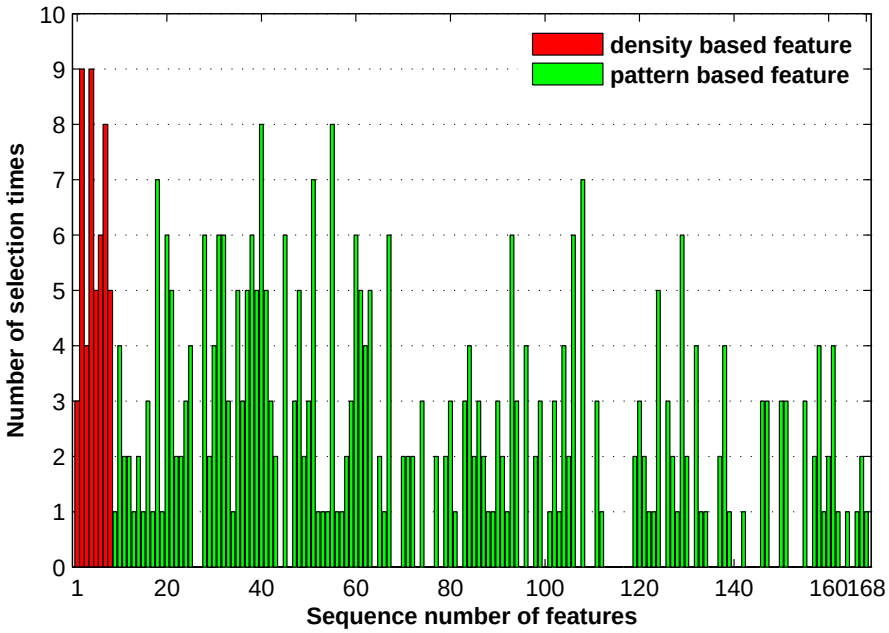


Figure 6.8: The overview of the selected features in the ten feature sets. The horizontal axis indicates the sequence number of features ranging from 1 to 168, and the vertical axis indicates the number of times each feature was selected over the five rounds.

the five rounds. It is shown that 20 features among the 168 features were commonly selected (the number of times which were selected was more than 5) in the ten feature sets: 4 from the density related features (50% of 8) and 16 from the pattern related features (10% of 160).

Subsequently, we performed two-fold cross-validation based on the reduced feature space. The classification results can be found in Table 6.12, including the classification accuracy for each fold and the overall classification accuracy for each round. The average overall classification accuracy was 71.15% and the standard deviation was 0.87% over the five rounds and 2.27% over the ten classifer models. This indicated a slightly improved performance than using the full feature space. It should be noted that the number of components in the combined model of breast tissue was reduced to less than one fourth of its former length. The corresponding confusion matrices are shown in Table 6.11 (right column). For each BIRADS class, there was no significant improvement in the classification accuracy over the five confusion matrices. The total percentages improved over the five rounds were 2%, 5%, 11%, and 5% for the four BIRADS classes, respectively.

Table 6.12: Classification results when using the selected feature sets, including the classification accuracy for each fold in the two-fold cross-validation and the overall classification accuracy for each round.

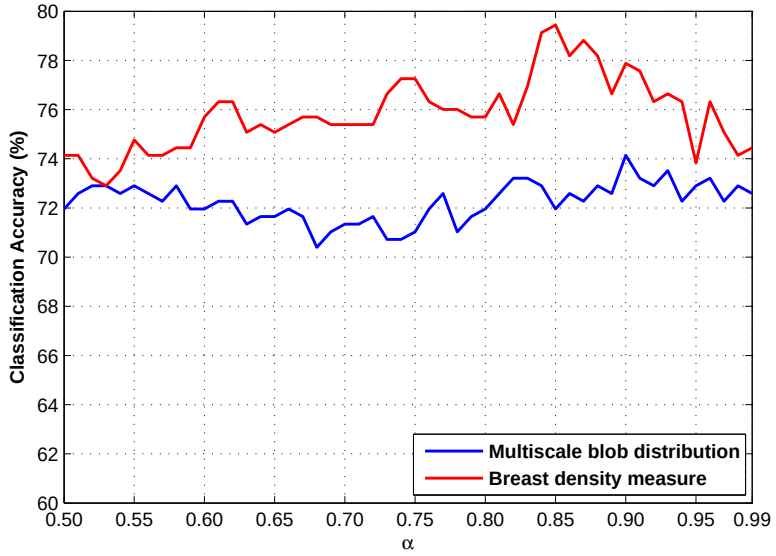
Test Data	Round 1	Round 2	Round 3	Round 4	Round 5
Fold I	71.43%	72.67%	68.94%	74.53%	70.19%
Fold II	72.50%	68.13%	71.25%	68.13%	73.75%
Overall	71.96%	70.40%	70.09%	71.34%	71.96%

6.3.4 Evaluation of Blob Based Representation

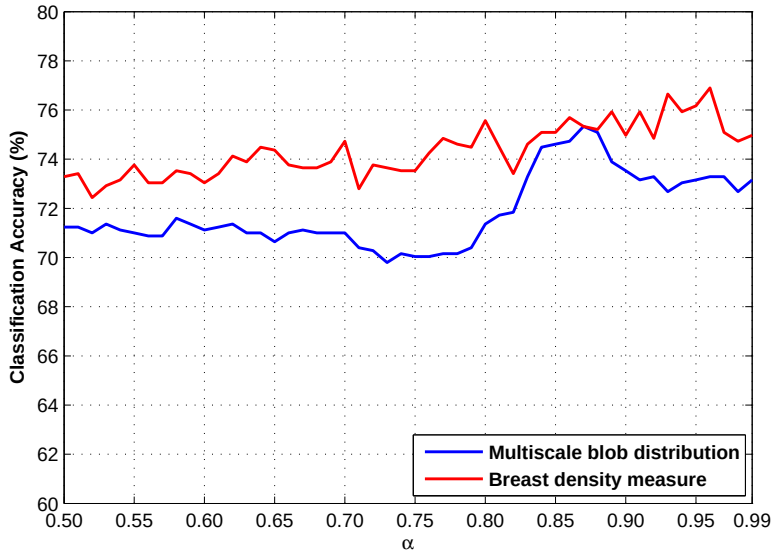
In Section 5.3, we presented a blob based representation of mammographic parenchymal patterns, which focused on the approximately blob-like tissue patterns. Multiscale blobs were detected by searching for the local extrema in the Laplacian scale-space representation of mammographic images over a range of scales. Qualitative relations among the resulting blobs were incorporated to merge the closely located blobs and as such a concise outline of dense tissue was retained. As described in Section 5.3.3, the distribution of the blobs over multiple scales can provide a quantitative description of breast density. A measure of breast density ($BLOB$) was defined on the basis of the multiscale blob representation which can be used for mammographic risk assessment.

The full MIAS database and the large subset of the DDSM database (we only used the MLO view here) were used for evaluating the validity of the proposed blob based representation for breast density classification and its potential for mammographic risk assessment. The parameters were set to the same values as in Section 5.3. Blobs were detected at 10 scales where the first scale was 8 pixels and the scale factor was $\sqrt{2}$. We investigated the performance of the 10-dimensional blob representation and the 1-dimensional breast density measure, respectively. For the 10-dimensional blob representation, the distribution of blobs over 10 scales was used as the feature vector in which each element N_i, representing the number of blobs at scale i, was normalised by the factor A_{blob_i}/A_{breast} where A_{blob_i} represents the area of blobs at scale i and A_{breast} represents the area of the breast. For the 1-dimensional breast density measure ($BLOB$), the normalised value ($BLOB_{norm}$) was used. The kNN based classifier was employed where the dissimilarity between the test object and the training sample was measured using the χ^2 distance. The leave-one-woman-out and leave-one-image-out cross-validation methodologies were used for the MIAS and DDSM databases, respectively.

As described in Section 5.3.2, there is an important parameter α in the step of blob merging, which controls the overlapping extent among the blobs after merging. Figure 6.9 shows the overall classification accuracy of the BIRADS density classification with respect to different values of α varying from 0.50 to 0.99. It



(a) MIAS



(b) DDSM

Figure 6.9: The overall classification accuracy with respect to the parameter α in the step of blob merging: (a) the MIAS database; (b) the DDSM database. The blue and red curves indicate the results obtained using the 10-dimensional multiscale blob distribution and the 1-dimensional breast density measure, respectively.

is shown that the breast density measure ($BLOB_{norm}$) obtained superior performance to that obtained using the multiscale blob distribution. This might be explained by the fact that the mammograms belonging to the same BIRADS class could have the large morphological variability in the blob-like dense tissue regions, which results in the variance of blob distributions. Specifically, two breasts having a similar percent density might have significantly different blob distributions

Table 6.13: Confusion matrices for the MIAS database: (a) using the multiscale blob distribution; (b) using the defined breast density measure.

(a) Blob distribution (CA = 74.14%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	77	10	0	0	89%
	II	19	79	5	0	77%
	III	1	18	68	7	72%
	IV	0	3	20	14	38%

(b) $BLOB_{norm}$ (CA = 79.44%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	72	15	0	0	83%
	II	5	88	10	0	85%
	III	0	12	74	8	79%
	IV	0	0	16	21	57%

(multiple blob-like dense regions detected over a series of small scales may indicate an equivalent percent density with only one blob-like dense region detected at a large scale). Transforming the multidimensional distribution of blobs into a single measure of breast density can reduce this kind of variability. For the MIAS database, the maximum classification accuracy was 74.14% when $\alpha = 0.9$ by using the multiscale blob distribution, and $BLOB_{norm}$ obtained the maximum classification accuracy of 79.44% when $\alpha = 0.85$. For the DDSM database, the maximum classification accuracy was 75.33% ($\alpha = 0.87$) and 76.90% ($\alpha = 0.96$) for the multiscale blob distribution and $BLOB_{norm}$, respectively. It seems that using a large value for α in the range of $[0.85, 1]$ tends to obtain better results, which is reasonable in the sense that only very closely located blobs are removed that more possibly derive from a single blob-like structure in the breast. The best classification accuracy for the DDSM database was slightly lower than that for the MIAS database. Similarly, this might be due to the fact that the DDSM database used here contained a larger number of mammograms than the MIAS database, which might introduce higher variance of breast tissue patterns.

Tables 6.13 and 6.14 show the resulting confusion matrices of the multiscale blob distribution and the defined breast density measure for the MIAS database and the DDSM database, respectively. When focusing on each BIRADS class classification, the best performance was obtained by BIRADS I and II; BIRADS III showed slightly worse performance when compared to BIRADS I and II; while BIRADS IV consistently indicated the worst performance with a large number of mammograms being misclassified into BIRADS III. This might be partially explained by the unbalanced distribution of these two neighbouring classes in the training set. Another reason might be that the extremely dense breasts in BIRADS IV have a more homogeneous tissue appearance all over the breast region, which possibly increases the incidence of false negatives due to the less-structured dense tissue.

Furthermore, evaluation based on the stratified 10-fold cross-validation (10-FCV) was performed for the DDSM database. Similarly, ten rounds of 10-FCV were run

Table 6.14: When using the leave-one-woman-out cross-validation methodology for evaluation, the resulting confusion matrices for the DDSM database: (a) using the multiscale blob distribution; (b) using the defined breast density measure.

(a) Blob distribution (CA = 75.33%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	88	18	0	0	83%
	II	34	276	25	1	82%
	III	1	46	192	16	75%
	IV	0	6	58	70	52%

(b) $BLOB_{norm}$ (CA = 76.90%)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	83	23	0	0	78%
	II	19	290	27	0	86%
	III	0	53	180	22	71%
	IV	0	5	43	86	64%

Table 6.15: When using the stratified 10-fold cross-validation methodology for evaluation, the average confusion matrices for the DDSM database: (a) using the multiscale blob distribution; (b) using the defined breast density measure.

(a) Blob distribution (CA = $75.09 \pm 0.58\%$)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	88.4±1.7	17.6±1.7	0.0±0.0	0.0±0.0	83.4±1.6%
	II	32.6±2.6	277.0±2.5	25.8±1.2	0.6±0.5	82.4±0.7%
	III	1.0±0.0	48.5±1.4	187.3±2.6	18.2±2.0	73.5±1.0%
	IV	0.0±0.0	5.9±0.6	56.8±2.7	71.3±2.8	53.2±2.1%

(b) $BLOB_{norm}$ (CA = $76.15 \pm 0.23\%$)

		Automatic				
	BIRADS	I	II	III	IV	CA
Truth	I	83.4±1.3	22.6±1.3	0.0±0.0	0.0±0.0	78.7±1.2%
	II	21.3±2.2	285.0±2.4	29.7±1.5	0.0±0.0	84.8±0.7%
	III	0.0±0.0	53.5±2.0	177.7±4.0	23.8±2.9	69.7±1.6%
	IV	0.0±0.0	5.2±0.6	42.1±3.5	86.7±3.3	64.7±2.5%

to avoid bias due to the random partition. Using the 10-dimensional multiscale blob distribution and the breast density measure of $BLOB_{norm}$ for classification, the resulting average classification accuracies over the ten rounds of 10-FCV were $75.09 \pm 0.58\%$ ($\alpha = 0.87$) and 76.05 ± 0.23 ($\alpha = 0.96$), respectively. The standard deviations over the 100 classifier models were 4.55% and 4.75%. Table 6.15 shows the average confusion matrices over the ten rounds of 10-FCV. Very similar results were obtained compared to those of using leave-one-woman-out.

In addition, we investigated the linear correlation of the defined breast density measure between the left and right MLO mammograms from the same woman in the MIAS database. Figure 6.10 shows the scatter plots of $BLOB_{norm}$ of 160 pairs of mammograms (mdb295ll and mdb296rl were excluded) in the MIAS database. A strong positive correlation between both sides (Pearson $r = 0.85$, Spearman

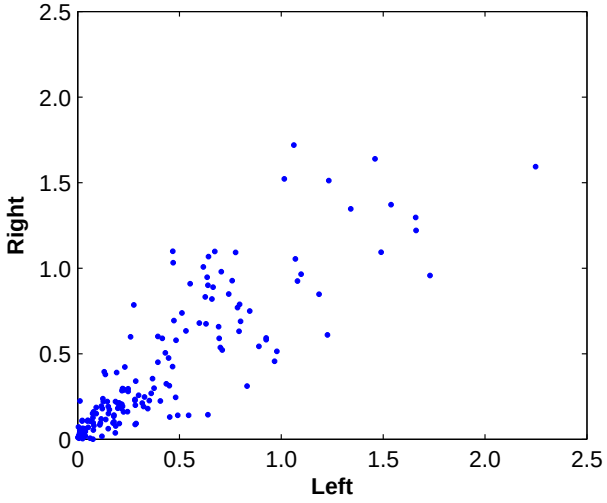


Figure 6.10: Scatter plots of $BLOB_{norm}$ of 160 pairs of bilateral mammograms in the MIAS database.

$rho = 0.87$) was indicated, which was similar with that obtained using the breast density measure defined based on the topographic approach (Pearson $r = 0.85$, Spearman $rho = 0.85$).

6.4 Summary

In this chapter, we evaluated all the methods proposed in the previous chapters for the purpose of mammographic risk assessment. At the beginning of the chapter, the data used in the experiments was introduced, which was taken from two benchmark databases in the studies of mammographic images. Subsequently, the evaluation methodologies used in this chapter were presented, including the classification algorithm based on the kNN classifier, the cross-validation schemes, and how to represent and analyse the classification results in the form of confusion matrices. On the basis of this, the proposed methods were evaluated in sequence. Firstly, the two breast density segmentation methods presented in Chapter 3 were evaluated. The MFCM based method was evaluated using the MIAS database. The best classification accuracies were $66.95 \pm 0.63\%$ and $67.60 \pm 1.25\%$ for the two groups of training sets (Train I and Train II), respectively. The topographic representation based method was evaluated using both the MIAS and DDSM databases. When using the leave-one-woman-out cross-validation methodology for evaluation, the resulting classification accuracy was 76.01%, 75.21%, and 74.61% for the MIAS database, the MLO view dataset of the DDSM database, and the CC view dataset of the DDSM database, respectively. This increased to 81.23% for the DDSM data when combining the MLO and CC views. For the DDSM

database, the stratified 10-fold cross-validation methodolgy was also employed for evaluation in addition to leave-one-woman out and equivalent classification results were indicated. The average classification accuracies over the ten 10-FCV rounds were $75.09 \pm 0.62\%$, $74.34 \pm 0.72\%$, and $80.88 \pm 0.44\%$ for the MLO view, the CC view, and the combination of the two views, respectively.

After that, the five strategies for breast tissue appearance modelling based on local features were evaluated using the MIAS database. The performance of the individual strategies was tested with respect to different settings of the parameters in these approaches and the correlation between every two approaches was investigated. The classification accuracies of the five strategies were 60.12%, 72.27%, 70.40%, 69.78%, and 74.77%, respectively. When combining the five classifier models built by the five modelling approaches, an classification accuracy of 77.88% was achieved. In addition, a combined model of breast tissue was generated by incorporating both density and tissue patterns. The dimensionality reduction of the resulting breast tissue model was also investigated using the SFS algorithm. The classification accuracies were $69.97 \pm 0.61\%$ and $71.15 \pm 0.87\%$ for the full dimensinality of the feature space and the reduced feature space, respectively.

Finally, the evaluation of the multiscale blob based representation of parenchymal patterns was presented. The MIAS database and the MLO view dataset of the DDSM database were used for evaluation. When using the leave-one-woman-out cross-validation methodology, the best classification accuracies were 79.44% and 76.90% obtained by the defined breast density measure $BLOB_{norm}$ for the MIAS and DDSM data, respectively. In addition, ten rounds stratified 10-fold cross-validation were performed for the DDSM data, the average classification accuracy over the ten rounds was $76.15 \pm 0.23\%$. A full comparison of our proposed methods with closely related publications and further discussions will be provided in Chapter 8.

Chapter 7

Topology Based Classification of Microcalcification Clusters

In the previous chapters, we have presented different approaches to analysing mammographic images for mammographic risk assessment, which focused on modelling breast tissue density/patterns in mammograms. In this chapter, we propose a novel method for the classification of microcalcification clusters in mammograms. The topology/connectivity of individual microcalcifications in a cluster is analysed using multiscale morphology. This is distinct from existing approaches which tend to concentrate on the morphology of individual microcalcifications and/or global (statistical) cluster features. A microcalcification graph is constructed to represent the topological structure of a cluster and its two properties associated with the connectivity form the basis for classification. This chapter is constructed as follows. Section 7.1 provides a general introduction to the difficulty in classifying microcalcification clusters as malignant or benign. Section 7.2 presents a brief review of the various methods for the characterisation and classification of microcalcifications in the literature. Two fundamental morphological operators, erosion and dilation, are described in Section 7.3, providing a basis introduction to mathematical morphology. Details of the data used in our experiments are described in Section 7.4. Section 7.5 presents the proposed method for the classification of microcalcification clusters. In Section 7.6, the validity of the proposed method is evaluated using two well-known digitised datasets and a full-field digital dataset. Four k-Nearest Neighbours based classifiers are employed and the achieved classification results including classification accuracy and receiver operating characteristic (ROC) analysis are presented. Finally, discussion and conclusions are provided in Section 7.7. A full comparison with related work reviewed in Section 7.2 is presented. In addition, highlights, limitations and future directions of the proposed method are discussed.

7.1 Introduction

As described in Section 1.4, it is difficult and time consuming for radiologists to interpret mammograms and distinguish malignant from benign microcalcifications. Figure 7.1 shows two examples of malignant and benign microcalcification clusters. The difficulty in the interpretation of microcalcifications results in a high rate of unnecessary biopsy examinations (Cheng *et al.*, 2003; Soltanian-Zadeh *et al.*, 2004). In order to improve the performance of radiologists, computer-aided diagnosis (CAD) systems have been applied to reduce the false positive rate while maintaining sensitivity (Cheng *et al.*, 2003; Papadopoulos *et al.*, 2005).
According to some studies on evaluation of breast microcalcifications, malignant microcalcifications tend to be small, numerous and densely distributed because they lie within the milk ducts and associated structures in the breast and follow the ductal anatomy. However, benign microcalcifications are generally larger, smaller in number and more diffusely distributed as these calcifications form within the breast stroma, benign cysts or benign masses (Sickles, 1986; Feig *et al.*, 1987). These differences result in variations in the distribution and closeness of microcalcifications within the clusters and provide radiologists with information which enables decisions regarding the need for further assessment and possible breast biopsy to be made. Hence we propose a hypothesis that microcalcification clusters can be classified as malignant or benign based on their topological structure.

7.2 Related Work

Numerous methods for the (semi-)automatic analysis of microcalcifications in mammograms have been proposed. A variety of features have been studied in the literature for the characterisation and classification of microcalcifications, such as shape, morphological, cluster and texture features (Cheng *et al.*, 2003).
Shen *et al.* (1994) developed three shape factors, including compactness, moments and Fourier descriptors, to quantitatively measure the roughness of individual microcalcifications, which were computed based on the extracted boundaries of microcalcifications. Similarly, Ma *et al.* (2010) proposed a novel shape metric to quantify the roughness of a microcalcification. A three level wavelet transform was used to estimate the frequency of the normalised distance signature of each microcalcification contour. Dengler *et al.* (1993) employed a morphological filter to reconstruct the original shape of smoothed microcalcifications. Betal *et al.* (1997) applied morphological operations to analyse four shape properties of segmented microcalcifications, which were infolding, elongation, narrow irregularities and wide irregularities. Chan *et al.* (1998) used morphological features to describe

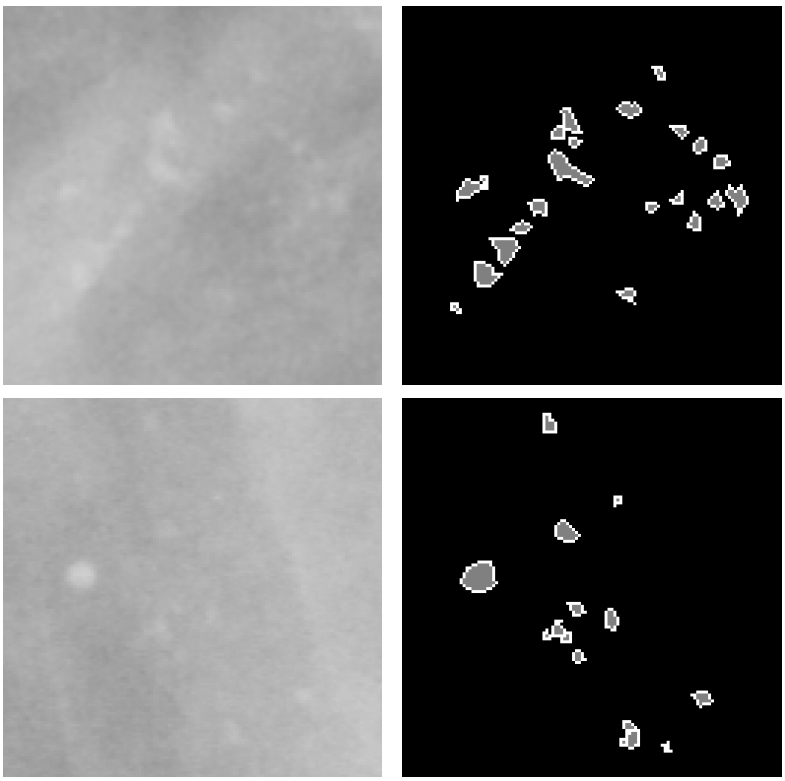


Figure 7.1: Example microcalcification clusters: malignant (top row) and benign (bottom row). Left column: mammographic image patches taken from the MIAS database; right column: manual annotations made by an expert.

the size, shape and contrast of individual microcalcifications and their variations within a cluster. Wirth *et al.* (2004) proposed an approach to enhancing the contrast of microcalcifications based on a top-hat morphological filter, where non-flat structuring elements were used to preserve the intensity surface (3D appearance) of the image. Halkiotis *et al.* (2007) applied a combination of morphological opening and reconstruction to mark potential microcalcifications as connected components in a binary image. Cluster features such as cluster area, number of microcalcifications, average and standard deviation of distances between microcalcifications were exploited in (Betal *et al.*, 1997; Dhawan *et al.*, 1996; Jiang *et al.*, 1997; Papadopoulos *et al.*, 2005) to describe the global properties of clusters. In addition to the features based on the geometric distribution of individual microcalcifications, a set of features based on the morphology of microcalcifications were studied in (Rana *et al.*, 2007). A morphologic feature was defined in (Shao *et al.*, 2011) for predicting the pathological classification (benign or malignant) and grading (four grades). Most of the features above were extracted based on the binary segmented microcalcifications and/or the cluster region, where greylevel or texture features were not taken into account. The greylevel variations and texture features were investigated for malignancy analysis in (Dhawan *et al.*, 1996; Chan *et al.*, 1997).

The global texture features were computed based on the spatial greylevel dependence (co-occurrence) matrix, and the local texture features were extracted based on the wavelet transform.
The performance of different types of features (Soltanian-Zadeh *et al.*, 2004) and methodologies (Ren, 2012) for microcalcification classification were compared. An image retrieval based approach was proposed in (Wei *et al.*, 2009), where similar cases stored in the database were retrieved as references to classify microcalcifications. An artificial neural network (ANN) classifier was employed in (Ren *et al.*, 2011), which included patient and image information.
For the classification of malignant and benign microcalcifications, the values of area under the ROC curve achieved by these publications range from 0.74 to 0.96. A complete comparison to these various approaches will be provided in Section 7.7.
There are also graph based approaches in histopatological image analysis (Gunduz *et al.*, 2004; Bilgin *et al.*, 2007, 2010). In these approaches, the cell graphs are generated based on the positional coordinates of the cells within the histopatological images, where the cells are represented as the nodes and the edges are formed based on the pairwise distance relationship between the cells. The graph theoretical features are extracted from the cell graphs to reflect their topological properties and the resulting features are subsequently used for classification of benign, invasive and noninvasive cancerous tissues.

7.3 Mathematical Morphology

Mathematical morphology (MM) is a theory for the analysis of spatial structures, which aims at analysing the shape and form of objects (Soille, 2003). It has been widely applied to image processing and analysis such as filtering, segmentation, classification, and texture analysis. The MM based approaches to image processing are achieved using morphological operators, which can extract relevant structures of the image by probing the image with a pre-defined shape called a structuring element (SE). The shape of the SE is chosen according to a prior knowledge on the geometric properties of objects to be processed. There are two fundamental morphological operators: erosion and dilation. All other morphological operators are composed of these two elementary operations.

7.3.1 Erosion

The erosion of a set X by a structuring element B, denoted by $\varepsilon_B(X)$, is defined as the locus of points $\mathbf{x}$: when the origin of B is placed at $\mathbf{x}$ (denoted by $B_\mathbf{x}$), it is included in X.

$$\varepsilon_B(X) = \{\mathbf{x} \mid B_\mathbf{x} \subseteq X\}. \tag{7.1}$$

The erosion of X by B can also be expressed with an intersection of set translations of X by the vectors $-\mathbf{b}$ of B:

$$\varepsilon_B(X) = \bigcap_{\mathbf{b} \in B} X_{-\mathbf{b}}. \tag{7.2}$$

This definition can be extended to binary and grey-scale images. The erosion of an image I by a structuring element B, denoted by $\varepsilon_B(I)$, is defined as the minimum of the translations of I by the vectors $-\mathbf{b}$ of B:

$$\varepsilon_B(I) = \bigwedge_{\mathbf{b} \in B} I_{-\mathbf{b}}. \tag{7.3}$$

Hence, the resulting value at a given pixel $\mathbf{x}$ in the eroded image $\varepsilon_B(I)$ is the minimum pixel value of the image I within the window defined by the structuring element B when its origin is at $\mathbf{x}$:

$$[\varepsilon_B(I)](\mathbf{x}) = \min_{\mathbf{b} \in B} I(\mathbf{x} + \mathbf{b}). \tag{7.4}$$

7.3.2 Dilation

The dilation of a set X by a structuring element B, denoted by $\delta_B(X)$, is defined as the locus of points $\mathbf{x}$: when the origin of B coincides with $\mathbf{x}$ (denoted by $B_\mathbf{x}$), it hits X.

$$\delta_B(X) = \{\mathbf{x} \mid B_\mathbf{x} \cap X \neq \emptyset\}. \tag{7.5}$$

Note that the dilation of a single pixel $\mathbf{x}$ outputs the reflected B centred at $\mathbf{x}$: $\delta_B(\mathbf{x}) = \check{B}_\mathbf{x}$. The dilation of X by B can also be expressed with a union of set translations of X by the vectors $-\mathbf{b}$ of B:

$$\delta_B(X) = \bigcup_{\mathbf{b} \in B} X_{-\mathbf{b}}. \tag{7.6}$$

This definition can be extended to binary and grey-scale images. The dilation of an image I by a structuring element B, denoted by $\delta_B(I)$, is defined as the maximum of the translations of I by the vectors $-\mathbf{b}$ of B:

$$\delta(I) = \bigvee_{\mathbf{b} \in B} I_{-\mathbf{b}}. \tag{7.7}$$

Therefore, the resulting value at a given pixel $\mathbf{x}$ in the dilated image $\delta_B(I)$ is the

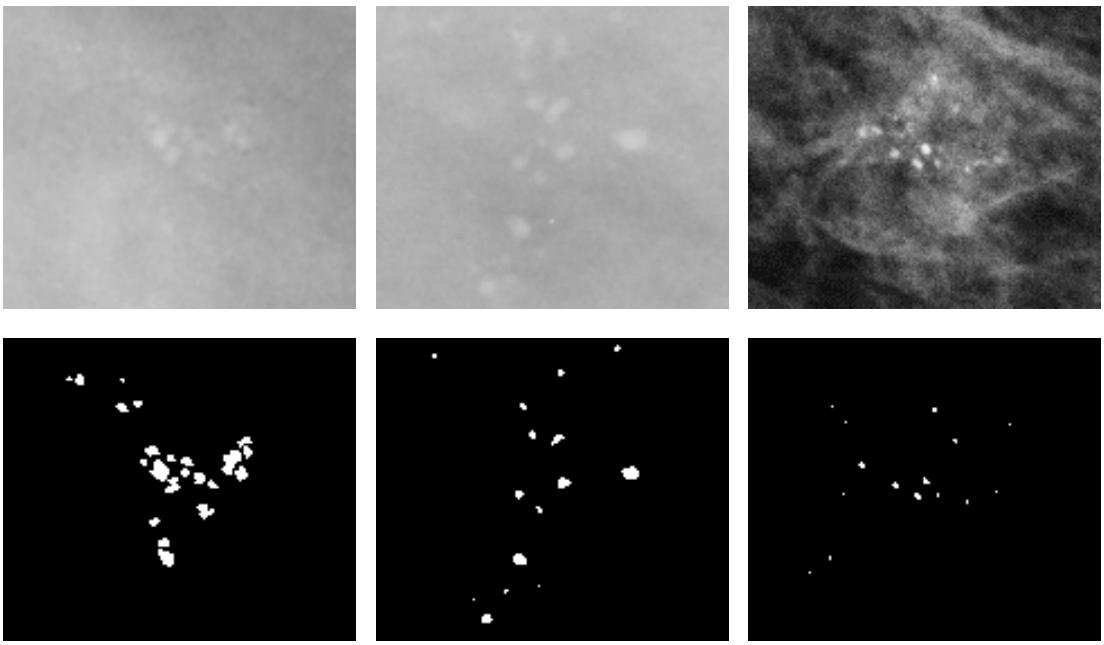


Figure 7.2: Example image patches and the corresponding automatic detection results selected from the three datasets: MIAS (left column), DDSM (middle column), and Digital (right column). The size of the three image patches is cropped to 157×135 pixels (images are zoomed to improve the visibility).

maximum pixel value of the image I within the window defined by the structuring element B when its origin is at $\mathbf{x}$:

$$[\delta_B(I)](\mathbf{x}) = \max_{\mathbf{b} \in B} I(\mathbf{x} + \mathbf{b}). \tag{7.8}$$

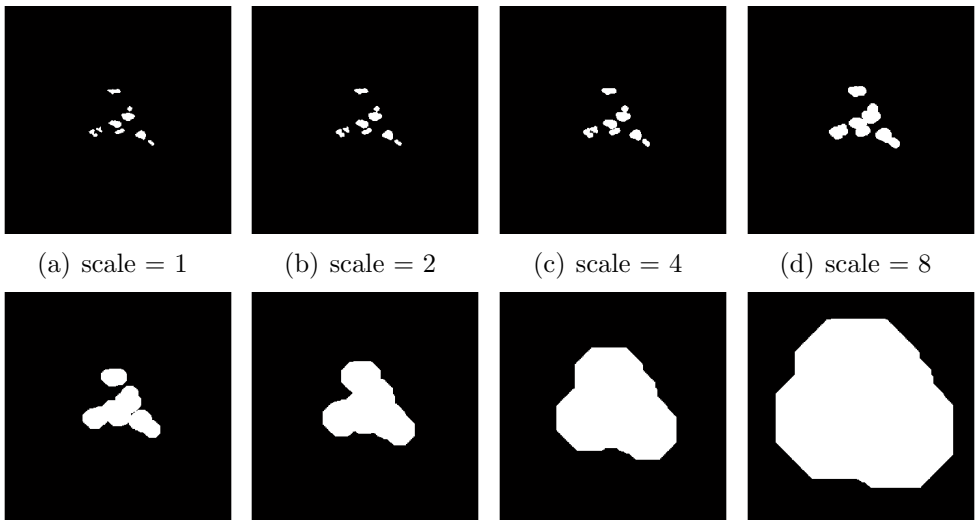
7.4 Data

The data used in the experiments consists of three datasets, containing image patches of different cases (taken from different mammograms). The first dataset was taken from the MIAS database, containing 20 image patches with the same size of 512×512 pixels. The spatial resolution is $50\mu m \times 50\mu m$ per pixel and quantised to 8 bits with a linear optical density in the range $0 - 3.2$. The second dataset was extracted from the DDSM database, containing 300 image patches with variable sizes, and their average size is 482×450 pixels. The mammograms were digitised by one of four scanners: DBA M2100 ImageClear (42 microns per pixel, 16 bits), Howtek 960 (43.5 microns per pixel, 12 bits), Lumisys 200 Laser (50 microns per pixel, 12 bits), and Howtek MultiRad850 (43.5 microns per pixel, 12 bits). In contrast to the first two datasets, the third dataset contains 25 full-field digital image patches extracted from a non-public mammographic database (referred to as Digital). The mammograms were acquired using a Hologic Selenia

mammograph, with a resolution of 70 microns per pixel and a depth of 12 bits. The size of these image patches also varies, and the average size is 352×301 pixels. Each image patch of the three datasets contains a microcalcification cluster. For all microcalcification clusters, mammographic diagnosis has been provided by biopsy: there are 9 malignant and 11 benign clusters in MIAS, 141 malignant and 159 benign clusters in DDSM, and 14 malignant and 11 benign clusters in Digital.
The proposed method works on binary detection/annotation images of microcalcifications, where 0s stand for "normal" tissue, and 1s represent microcalcifications. For automatic detection of microcalcifications, the approach developed by Oliver *et al.* (2012) is applied. A probability image is obtained for each image patch, in which high values indicate a high probability of microcalcifications being present. The probability image is binarised using an appropriate threshold in order to obtain a binary image. The threshold is determined using ROC analysis of the resulting microcalcifications. The highest sensitivity and specificity on the ROC curve are achieved by using this threshold. As the ground truth for ROC analysis, for the MIAS dataset, the precise location of individual microcalcifications was manually annotated by an expert (each microcalcification was labelled and segmented from the surrounding tissue (Figure 7.1)), while for the remaining two datasets, two experts annotated the location of microcalcification clusters instead of individual microcalcifications (each cluster was considered as a whole object) with ellipses circumscribing the clusters (each expert annotated a non-overlapping subset of image patches). For each image patch, a ROC curve is produced by varying the threshold between 0 and 1. The point on the ROC curve that is the closest to the point (0,1) (1 sensitivity and 1 specificity) is considered as the optimal detection, and the corresponding threshold value is selected to segment microcalcifications. Example image patches selected from the three datasets and the corresponding binary detection images are shown in Figure 7.2, showing significant variation in the original images and the automatic detection results (the original work by Oliver *et al.* (2012) showed better results for digital data when compared to digitised data).

7.5 Methodology

We analyse the topology of microcalcification clusters at multiple scales and define a multiscale topological feature vector to discriminate malignant from benign cases. We estimate the connectivity between individual microcalcifications within a cluster using dilation at multiple scales. At each scale a graph of microcalcifications is constructed to represent the topological structure of microcalcifications within clusters. Two properties of microcalcification graphs are defined, which



(a) scale = 1 (b) scale = 2 (c) scale = 4 (d) scale = 8

(e) scale = 16 (f) scale = 32 (g) scale = 64 (h) scale = 128

Figure 7.3: An example of the multiscale dilation: microcalcifications are dilated with a disk-shaped structuring element over a range of eight scales.

can provide distinguishing information for benign and malignant microcalcification clusters. The multiscale topological feature vector is produced by combining the microcalcification graph properties at multiple scales.

7.5.1 Morphological Operation

Firstly, morphological dilation is performed on individual microcalcifications using a disk-shaped structuring element. The dilation works on binary annotation/ detection images at multiple scales. Here, the scale corresponds to the radius of the structuring element, which is measured in pixels. Figure 7.3 shows an example of the dilation process, which indicates the multiscale morphological dilation adds neighbouring pixels to the boundaries of individual microcalcifications, resulting in changes in the connectivity between individual microcalcifications within clusters. To illustrate the topological structure and connectivity of microcalcifications with respect to malignant and benign clusters, the dilation results of the two example microcalcification clusters in Figure 7.1 are shown in Figure 7.4, where the radius of the structuring element is equal to six pixels. Each individual microcalcification is ordered with a sequential number and the boundaries of dilated microcalcifications are displayed using different colours. As indicated in Section 7.1, we can see from Figure 7.4 that the malignant cluster (top row) contains a larger number of microcalcifications which are close together, while the benign cluster (bottom row) contains fewer microcalcifications which are more distributed.

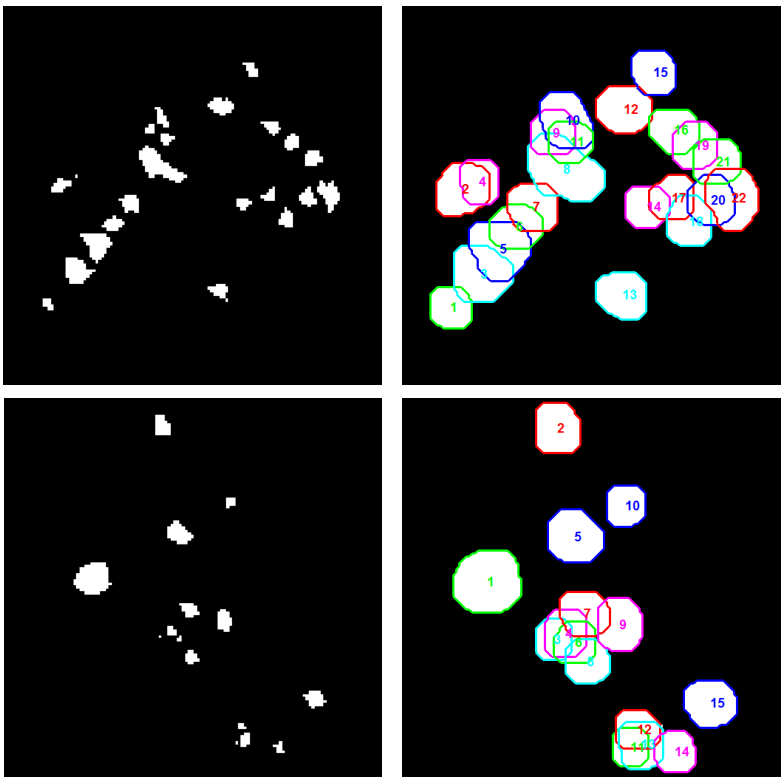


Figure 7.4: Morphological dilation of microcalcifications. Left column: binary annotation images of microcalcifications; right column: dilated microcalcifications using a disk-shaped structuring element with radius equal to six pixels. The microcalcification cluster regions are zoomed for better illustration. Microcalcification No. 16 for the benign case is not displayed as it falls outside the displayed region.

7.5.2 Microcalcification Graph

A graph is an abstract representation of a set of objects, which can provide a graphical description of the structure or distribution of the objects (Conte *et al.*, 2004). In this work, the topology of individual microcalcifications within a cluster is represented in graphical form. A microcalcifications graph is constructed based on the spatial connectivity relationship between microcalcifications. In the graph, each node represents an individual microcalcification, and an edge is added to link two nodes if the two corresponding microcalcifications are connected or overlap with each other in the image plane. Here, we generate a directed graph where the nodes are ordered according to the spatial location of the corresponding microcalcifications in the image patch (the nodes are ordered in a left-to-right and bottom-to-top direction in our work, but alternative directions provide the same performance for the subsequent processing), and two connected nodes are linked by a directed edge from the lower to the higher numbered node. The resulting graphs of dilated microcalcifications in Figure 7.4 are shown in Figure 7.5. The numbering of nodes is consistent with Figure 7.4.

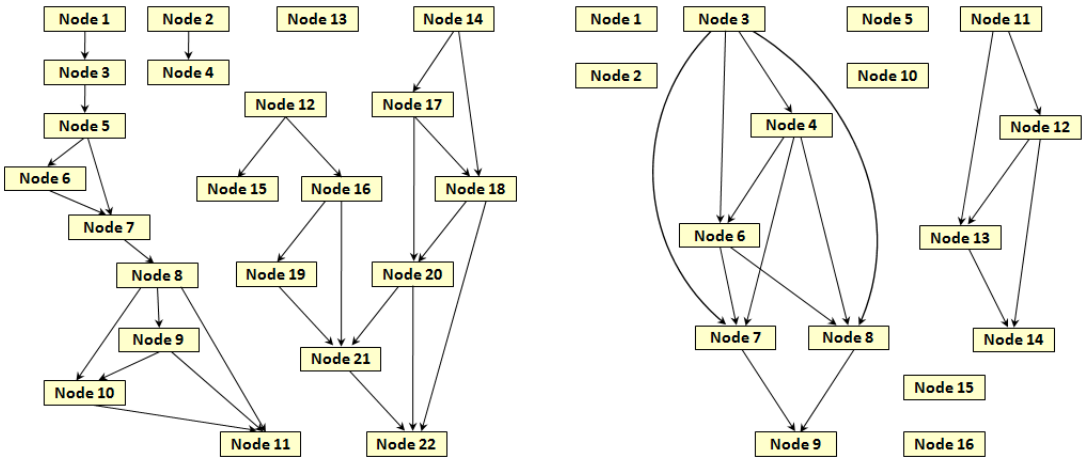
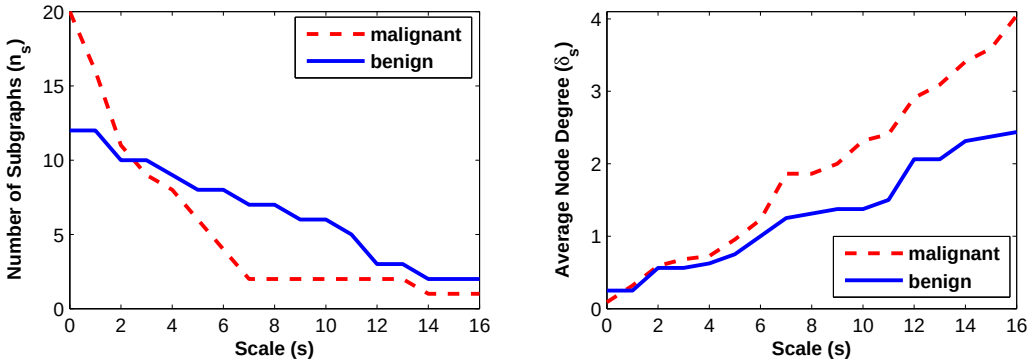


Figure 7.5: Microcalcifications graphs of example microcalcification clusters: (a) malignant ($n_s = 4$, $d_s = 1.23$); (b) benign ($n_s = 8$, $d_s = 1.00$). Note that the graphs are directed where the arrows indicate the direction of edges and the numbering of nodes is consistent with Figure 7.4

We define two microcalcifications graph properties. The first property is the number of independent subgraphs within the graph generated from one microcalcification cluster, which represents the number of independent connected components within the cluster. The second property is the average degree of all nodes within the directed graph. The degree of a node is defined as the number of edges starting from the node, which describes the connectivity of the corresponding microcalcification with its children.

For further analysis based on these two properties of microcalcification graphs, we give some basic definitions. We define an upper-triangular adjacency matrix to encode the microcalcification graph, denoted by $A = (a_{ij})$, $a_{ij} \in \{0, 1\}$, $i, j = 1, \ldots, m$, where m is the number of nodes within the graph. $a_{ij} = 1$ indicates node i and node j are connected, node i is the source node and node j is the sink node. A *source* node i is called a *root* node if $\sum_{k=1}^{m} a_{ki} = 0$. A *sink* node j is called a *terminal* node if $\sum_{k=1}^{m} a_{jk} = 0$. A *path* from node i to node j is defined as a sequence of nodes starting from node i and ending with node j. More definitions and propositions on graphs can be found in (Diestel, 2010).

The number of independent subgraphs (denoted by n) is determined by traversing the graph. We traverse the graph starting at each *root* node and explore as far as possible along each *path* until arriving at the *terminal* node. The traversal sequences including common nodes are combined into a single sequence. After combining the overlapping traversal sequences, all the resulting sequences are unique and independent. The number of these sequences is the number of independent subgraphs. The degree of node i (denoted by $d(i)$) is computed by $d(i) = \sum_{k=1}^{m} a_{ik}$,



(a) (b)

Figure 7.6: Multiscale topological features of example malignant and benign microcalcification clusters in Figure 7.1: (a) number of subgraphs; (b) average node degree. The features are extracted at 17 scales.

and the average node degree of m nodes is determined by $\delta = \frac{1}{m}\sum_{i=1}^{m} d(i)$.

7.5.3 Multiscale Topological Feature Vector

We construct a set of microcalcifications graphs $G = (G_0, G_1, \ldots, G_{S-1})$ based on morphologically dilated microcalcifications to represent the topology of microcalcification clusters at multiple scales, where S is the number of scales, and G_s denotes the microcalcification graph corresponding to scale s. We extract the two defined microcalcification graph properties from the resulting graph set G, which forms two vectors $N = (n_0, n_1, \ldots, n_{S-1})$ and $\Delta = (\delta_0, \delta_1, \ldots, \delta_{S-1})$, where n_s denotes the number of independent subgraphs at scale s, and δ_s denotes the average node degree at scale s. Finally, the two vectors are concatenated into a single feature vector, i.e. the multiscale topological feature vector, representing the multiscale topological characteristics of microcalcification clusters, which can be used for the classification of malignant and benign microcalcification clusters. The resulting N and D for the example microcalcification clusters in Figure 7.1 are shown in Figure 7.6, where the topology of microcalcification clusters is analysed at 17 scales ($s = 0, 1, \ldots, 16; S = 17$). The number of connected components within the malignant cluster decreases faster than the benign cluster as the dilation scale increases. The average node degree features of the malignant cluster are larger than those of the benign cluster.

7.6 Experimental Evaluation

7.6.1 Experimental Set-up

The k Nearest Neighbours (kNN) classifier (Duda *et al.*, 2001a) is a popular and conceptually intuitive instance-based learning approach. It does have a number of shortcomings, however, a number of alternatives are employed which attempt to address some of these. Fuzzy Nearest Neighbours (FNN) (Keller *et al.*, 1985) extends classical kNN by fuzzifying the memberships for test and training objects, thus allowing partial membership of classes. Fuzzy Rough Nearest Neighbours (FRNN) (Jensen & Cornelis, 2011; Mac Parthaláin *et al.*, 2010) models two different types of uncertainty: fuzziness (or vagueness) and indiscernibility (or lack of information). Vaguely Quantified Nearest Neighbours (VQNN) (Cornelis *et al.*, 2007) incorporates the uncertainty modelling of FRNN and also employs vague quantifiers which limit the influence that noisy data might have on the classification outcomes. These approaches offer further flexibility, improved generalisation, and retain human interpretability when compared to techniques such as neural networks and support vector machines (SVM). It should be noted that classical kNN is also employed for this evaluation such that the proposed approach can be easily compared with existing work in the literature.

The kNN classification was based on a simple majority vote, unless equal class probability was indicated, in which case a Euclidean weighted approach was used. For FNN, the default fuzzifier value of $m = 3.0$ was used in the experiments. FRNN is stable with respect to the value of k and returns similar results but slightly different models. VQNN on the other hand results in different models when the value for k is altered. A range of values for k were employed when generating the results in the following section.

To evaluate the performance of the classifier models built by using the multiscale topological feature vectors, a leave-one-out cross-validation (LOOCV) scheme was employed for all datasets, and an additional stratified 10 runs 10-fold cross-validation (10-FCV) scheme was employed for the DDSM dataset. Two evaluation metrics are used for this work. The first is overall classification accuracy (CA), which provides a summary of the performance for balanced datasets (such as the used datasets here). ROC analysis is used as the second evaluation approach, which can be employed in order to assess the predictive ability of a classifier, by using the Area-Under-the-ROC-Curve (A_z) metric (Beck & Schultz, 1986). True positive rate (TPR) is defined as the number of correctly classified malignant cases divided by the total number of malignant cases, and the false positive rate (FPR) is defined as the number of benign cases incorrectly classified as malignant di-

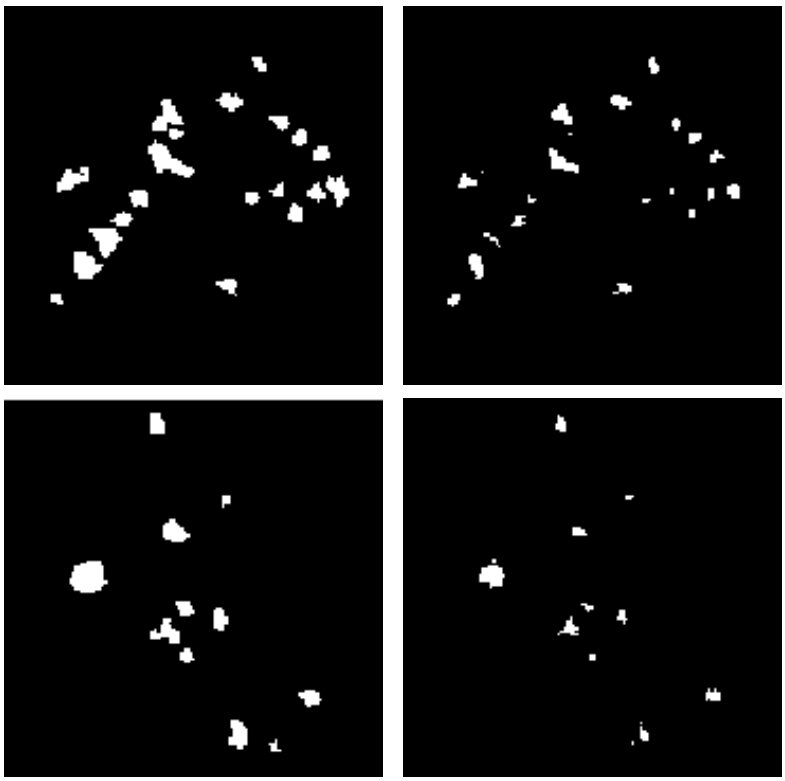


Figure 7.7: Manual annotations and automatic detection results of the example malignant (top row) and benign (bottom row) microcalcification clusters. Left column: manual annotations by an expert; right column: automatic detection results by Oliver *et al.* (2012).

vided by the total number of benign cases. All of the classification and evaluation aspects were completed using the Weka data mining suite (Hall *et al.*, 2009).
In order to offer a comparison between the classification results based on manually and automatically segmented microcalcifications (and also to investigate the robustness of the proposed method to microcalcification segmentation variations), it was tested using both manual annotations and automatic detection results for the MIAS dataset. As mentioned in Section 7.4, an expert manually annotated individual microcalcifications for all the image patches in the MIAS dataset. Figure 7.7 shows both manual annotations and automatic detection results of the example microcalcification clusters in Figure 7.1. It appears that the automatic detection approach tends to under-segment individual microcalcifications, such that the pixels close to the boundaries of individual microcalcifications are lost.

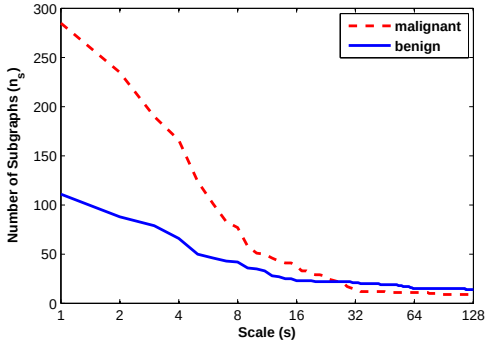
7.6.2 Experimental Results

In order to evaluate the potential of the proposed method in discriminating between malignant and benign microcalcification clusters, it has been tested using two digitised and one full-field digital datasets: MIAS, DDSM and Digital

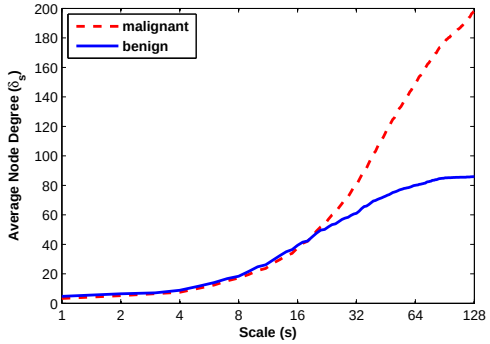
(see Section 7.4 for details). We have investigated a range of scales (S) defining the feature space. The dimensionality of the feature space is determined by S, as microcalcification graphs (and the extracted features) are generated at scales $s = 0, 1, \ldots, S-1$. The maximum scale used in the experiments was equal to 128, i.e. $S = 129$, and therefore the maximum dimensionality of the feature space was 258 (we extracted two features, n_s and d_s, from each graph generated at scale s). Figure 7.8 shows the overall distribution of the two features with respect to the scale for the malignant and benign cases of the four test datasets (two test datasets for the MIAS dataset: using manual annotation/automatic detection). As can be seen from the left column of Figure 7.8, the number of components within the malignant clusters is larger than the benign clusters at the beginning of the multiscale dilation (corresponding to smaller scales), and it decreases faster than that of the benign clusters as the scale increases. On the other hand, as shown in the right column of Figure 7.8, the average node degree of the malignant clusters is slightly smaller than that of the benign clusters at smaller scales, and it considerably increases when dilating microcalcifications with larger scales.

As a preliminary test, we used the Nearest Neighbour (1NN) classier and the LOOCV scheme to investigate the classification performance of the two features. Figure 7.9 shows the classification accuracy for three different combinations of the two features: using only the number of independent subgraphs, using only the average node degree, and using the combination of both features, which are referred to as *feature*1, *feature*2 and *feature*1+*feature*2, respectively. It is shown that *feature*1 obtains the highest classification accuracy in the lower part of the scale range and maintains stable performance over the higher part. This can be explained by the evidence observed in Figure 7.8 (left column) that *feature*1 indicates discriminative values for the malignant and benign clusters at smaller scales and becomes less discriminative when the scale increases to larger values (as the number of components within clusters stops decreasing when all close microcalcifications are merged into one object). In contrast, *feature*2 achieves its best performance in the higher part of the scale range. As shown in Figure 7.8 (right column), it provides more discriminative information for differentiating malignant from benign at larger scales. Therefore, it is considered that complementary information can be obtained by combining these two features over the whole range of scales. This has been indicated by Figures 7.9(a) and 7.9(d) in which better performance is achieved by using *feature*1+*feature*2. The integrated vectors of these two features, i.e. multiscale topological feature vectors as described in Section 7.5, were used for the subsequent experiments.

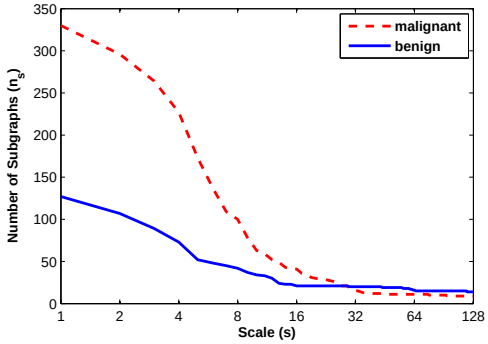
Figure 7.10 shows the classification results for the three datasets covering 129 scales. For the MIAS dataset, when using the manual annotations, the best CA



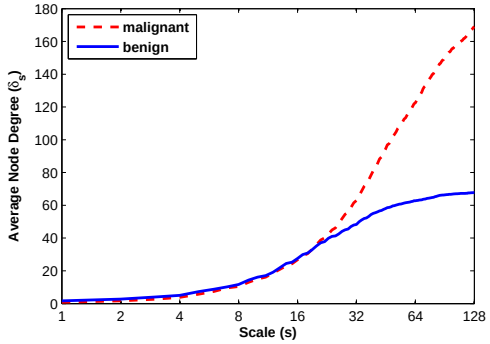
(a) n_s for MIAS (manual annotation)



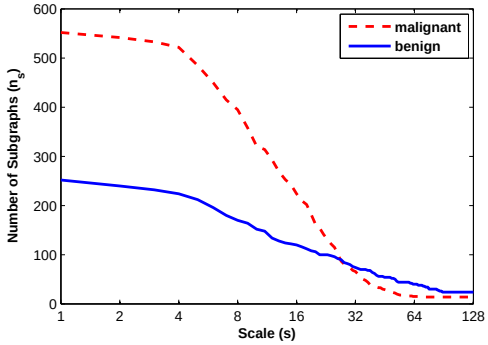
(b) δ_s for MIAS (manual annotation)



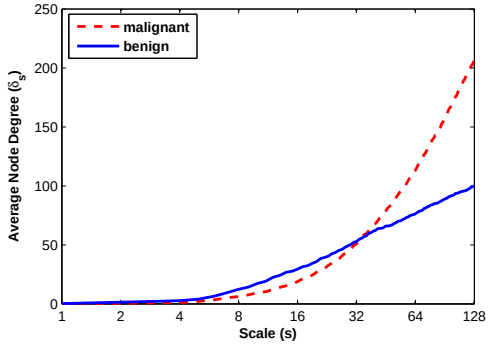
(c) n_s for MIAS (automatic detection)



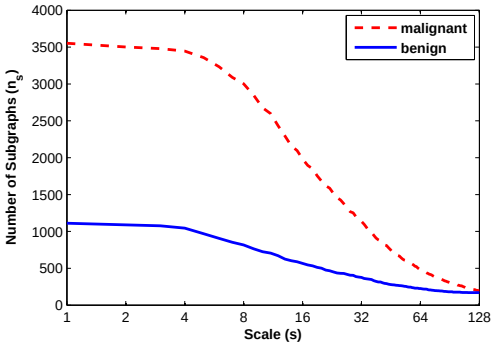
(d) δ_s for MIAS (automatic detection)



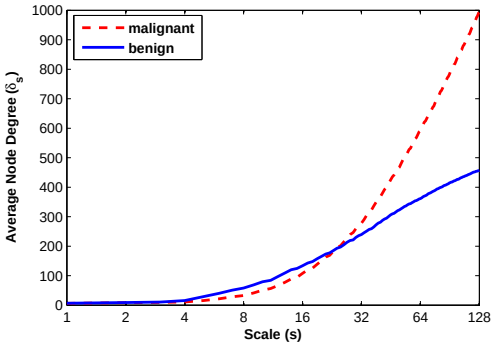
(e) n_s for Digital



(f) δ_s for Digital

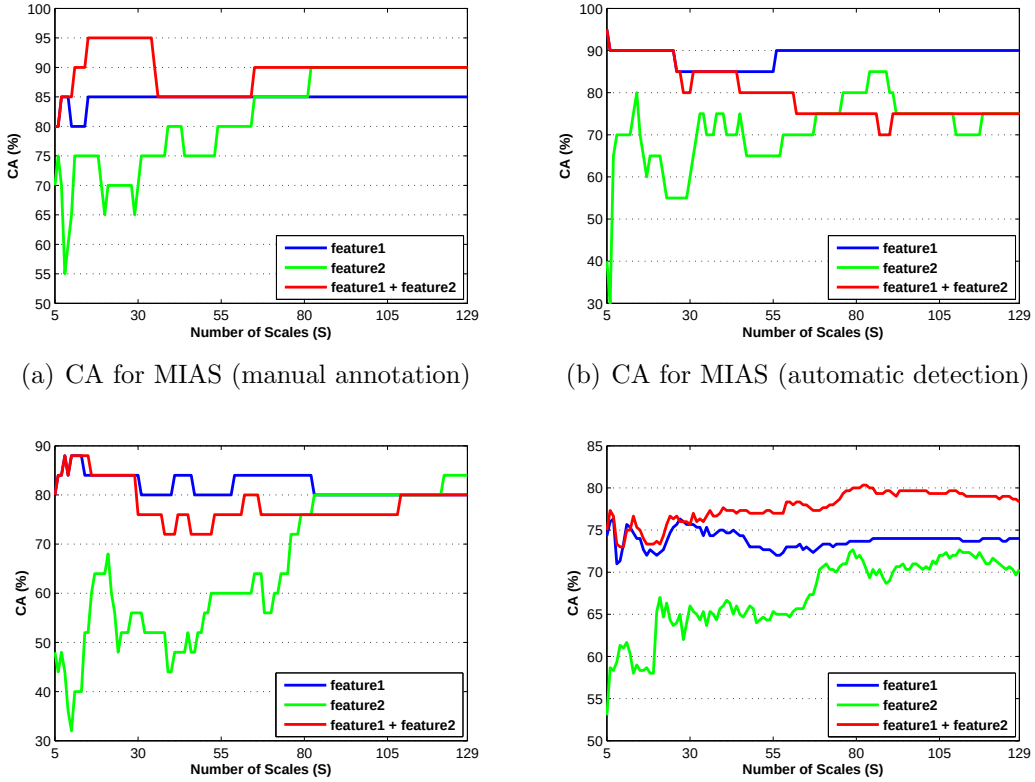


(g) n_s for DDSM



(h) δ_s for DDSM

Figure 7.8: The overall distribution of the defined topological features with respect to the scale for the malignant and benign microcalcification clusters of the test datasets. Left column: number of independent subgraphs (n_s) with respect to the scale (s); right column: average node degree (δ_s) with respect to the scale (s).



(a) CA for MIAS (manual annotation)

(b) CA for MIAS (automatic detection)

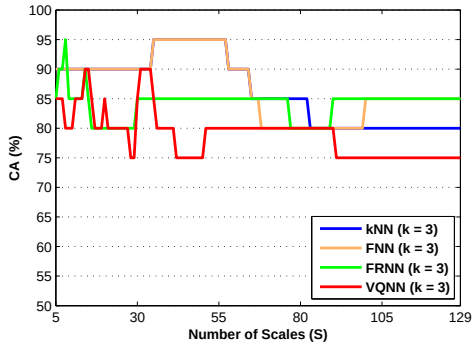
(c) CA for Digital

(d) CA for DDSM (LOOCV)

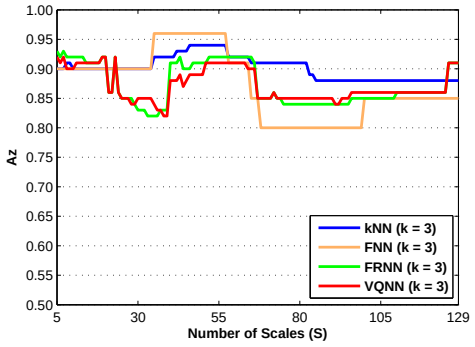
Figure 7.9: The classification accuracy obtained by using three different combinations of the two features. *feature*1: using only the number of independent subgraphs; *feature*2: using only the average node degree; and *feature*1+*feature*2: using the combination of both features.

was 95% with one benign case misclassified, produced by kNN, FRNN and FNN, and the largest A_z was 0.96 obtained by FNN; when using the automatic detection results, the best CA was also 95% obtained by FNN, with the same benign case misclassified, and the largest A_z of 0.96 was obtained by FNN and FRNN. For the Digital dataset, the best CA was 96% achieved by FRNN, with one malignant case misclassified, and the best A_z of 0.97 was also obtained by FRNN. For the DDSM dataset, when using LOOCV, kNN obtained the best CA of 84.3% and the largest A_z was 0.89 produced by kNN and VQNN; when using 10-FCV, kNN indicated the best performance, the obtained CA and A_z were $83.8 \pm 5.7\%$ and 0.90 ± 0.06, respectively (standard deviations were calculated across 100 classifier models (10 folds $\times$ 10 runs)). Table 7.1 shows a summary of the best classification results achieved by the four classifiers over 129 scales.

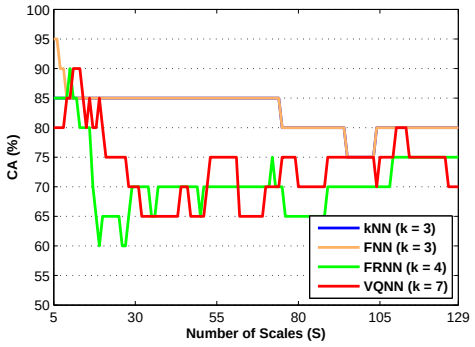
A statistical comparison of the classifiers was also made. The classification accuracy obtained over 129 scales (from scale 0 to scale 128) were aggregated for comparison. A paired t-test with significance level of 0.05 has been performed (over 10 runs for 10-FCV). The results obtained by the kNN classifier were used



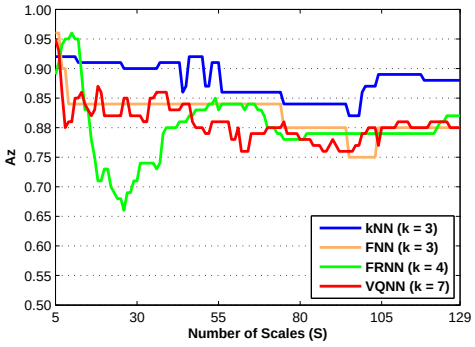
(a) CA for MIAS (manual annotation)



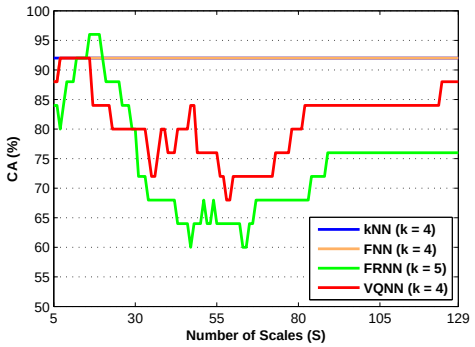
(b) Az for MIAS (manual annotation)



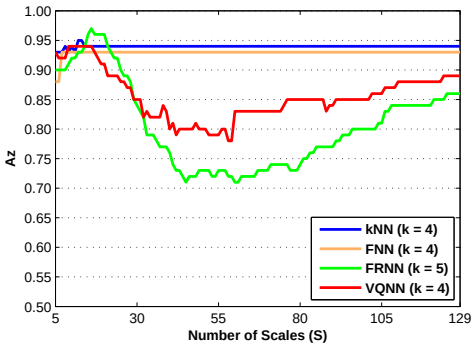
(c) CA for MIAS (automatic detection)



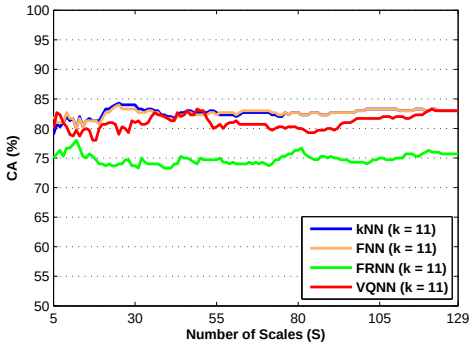
(d) Az for MIAS (automatic detection)



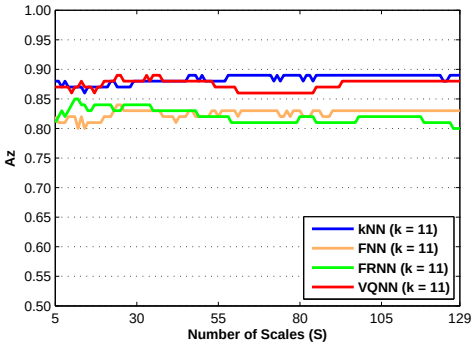
(e) CA for Digital



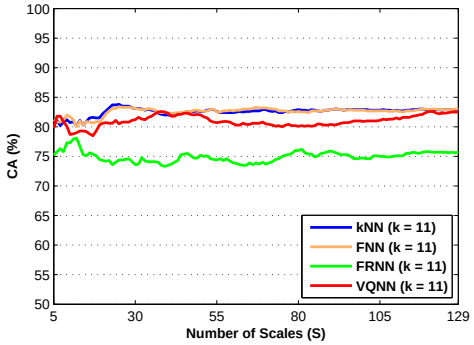
(f) Az for Digital



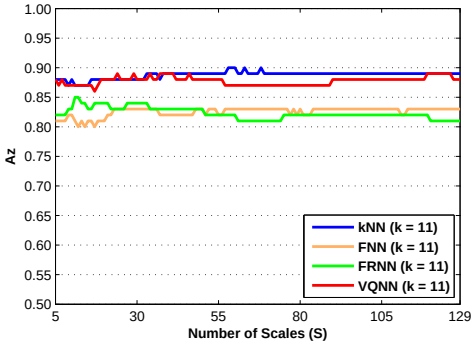
(g) CA for DDSM (LOOCV)



(h) Az for DDSM (LOOCV)



(i) CA for DDSM (10-fold CV)



(j) Az for DDSM (10-fold CV)

Figure 7.10: The classification results of the four classifiers for different datasets. Left column: classification accuracy (CA) with respect to different scale ranges; right column: Area-Under-the-ROC-Curve (A_z) with respect to different scale ranges. Note that some curves for different classifiers are overlapped.

Table 7.1: The best classification results of the four classifiers for different datasets over 129 scales. For 10-FCV, the results contain means and standard deviations resulting from 100 classifier models (10 folds $\times$ 10 runs).

(a) Classification Accuracy (%)

Test Data	kNN	FNN	FRNN	VQNN
MIAS (manual annotation)	95 ($k = 3$)	95 ($k = 3$)	95 ($k = 3$)	90 ($k = 3$)
MIAS (automatic detection)	85 ($k = 3$)	95 ($k = 3$)	90 ($k = 4$)	90 ($k = 7$)
Digital	92 ($k = 4$)	92 ($k = 4$)	96 ($k = 5$)	92 ($k = 4$)
DDSM (LOOCV)	84.3 ($k = 11$)	84.0 ($k = 11$)	78.0 ($k = 11$)	83.3 ($k = 11$)
DDSM (10-FCV)	83.8 ± 5.7 ($k = 11$)	83.4 ± 6.7 ($k = 11$)	78.1 ± 7.6 ($k = 11$)	82.6 ± 7.0 ($k = 11$)

(b) Area-Under-the-ROC-Curve

Test Data	kNN	FNN	FRNN	VQNN
MIAS (manual annotation)	0.94 ($k = 3$)	0.96 ($k = 3$)	0.93 ($k = 3$)	0.93 ($k = 3$)
MIAS (automatic detection)	0.92 ($k = 3$)	0.96 ($k = 3$)	0.96 ($k = 4$)	0.95 ($k = 7$)
Digital	0.95 ($k = 4$)	0.93 ($k = 4$)	0.97 ($k = 5$)	0.94 ($k = 4$)
DDSM (LOOCV)	0.89 ($k = 11$)	0.84 ($k = 11$)	0.85 ($k = 11$)	0.89 ($k = 11$)
DDSM (10-FCV)	0.90 ± 0.06 ($k = 11$)	0.83 ± 0.07 ($k = 11$)	0.85 ± 0.07 ($k = 11$)	0.89 ± 0.06 ($k = 11$)

as the baseline references for the paired t-test. The statistical significance results are summarised in Table 7.2. The three numbers in the table cell indicate a count of the number of times that the approach performs statistically better than, the same as, or worse than the baseline approach, respectively. For example, "0/107/22" indicates the results of this approach were statistically better than those of kNN for 0 times, equivalent to those of kNN for 107 times, and worse than those of kNN for 22 times. As can be seen from Table 7.2, kNN, FNN and VQNN performed equivalently for all the datasets. The results achieved by FRNN were statistically comparable with kNN, FNN and VQNN for the MIAS and Digital datasets (slightly worse for the Digital dataset), but were the statistically worst for the DDSM dataset.

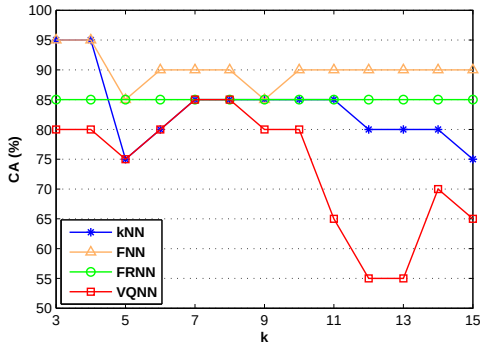
Table 7.2: The statistical comparison of the four classifiers based on the results obtained over 129 scales. For 10-FCV, the comparison was performed based on the results generated over 10 runs.

Test Data	kNN	FNN	FRNN	VQNN
MIAS (manual annotation)	0/129/0 ($k = 3$)	0/129/0 ($k = 3$)	0/129/0 ($k = 3$)	0/129/0 ($k = 3$)
MIAS (automatic detection)	0/129/0 ($k = 3$)	0/129/0 ($k = 3$)	0/129/0 ($k = 4$)	0/129/0 ($k = 7$)
Digital	0/129/0 ($k = 4$)	0/129/0 ($k = 4$)	0/107/22 ($k = 5$)	0/129/0 ($k = 4$)
DDSM (LOOCV)	0/129/0 ($k = 11$)	0/129/0 ($k = 11$)	0/18/111 ($k = 11$)	0/128/1 ($k = 11$)
DDSM (10-FCV)	0/129/0 ($k = 11$)	0/129/0 ($k = 11$)	0/10/119 ($k = 11$)	0/127/2 ($k = 11$)

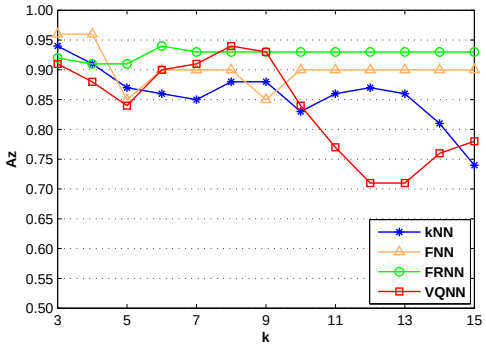
In addition, we have investigated a range of the number of neighbours (k) used in the classifiers. The k values used above (see Figure 7.10) were selected experimentally. Figure 7.11 shows the classification results of the four classifiers with respect to variation in the value of k for the three datasets. A fixed scale range was used for each test dataset. The number of scales (S) was set to 56, 6, 16, 58 and 58 for MIAS (manual annotation), MIAS (automatic detection), Digital, DDSM (LOOCV) and DDSM (10-FCV), respectively. The value of k was varied from 3 to 15 for each classifier. It should be noted from Figure 7.11 that the results were robust with regard to changes in the value of k. For the two smaller datasets, MIAS and Digital, kNN, FNN and VQNN produced slightly better performance when using smaller k values; while for DDSM, there was a slightly increasing trend when using larger k values. FRNN produced the most stable performance among all the classifiers for different k values.

7.7 Discussion and Conclusions

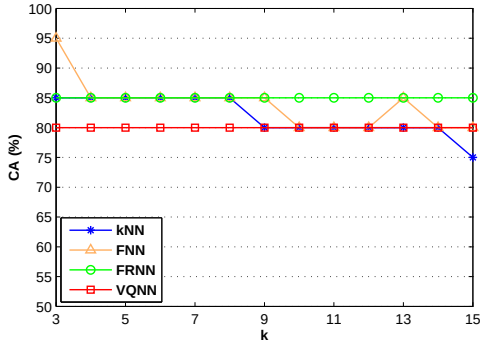
We have presented a method for classifying microcalcification clusters in mammograms based on morphological topology analysis. This is a novel approach to analyse microcalcifications in terms of the connectivity and topology for discriminating malignant from benign clusters. Unlike most features (e.g. shape) in previous publications extracted at a single scale, a representation of microcalcification clusters covering the multiscale characteristics was developed in this chapter. The topology/connectivity of microcalcification clusters was analysed using multiscale morphology. A set of microcalcification graphs were constructed to describe the topological structure of microcalcifications at multiple scales. When analysing the topology of microcalcification clusters, we focused on two microcalcification graph properties, the number of independent subgraphs and the degree of nodes. These two features can provide complementary information in describing the topological structure of microcalcification clusters, though they are both related to relative closeness of individual microcalcifications (better performance was indicated in the



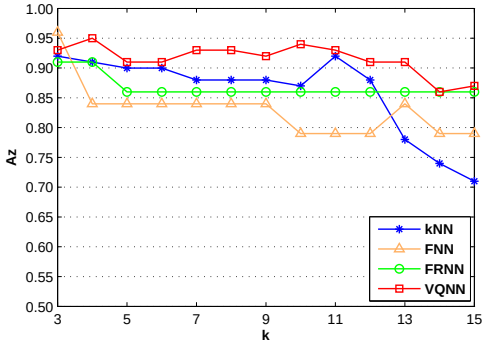
(a) CA for MIAS (manual annotation), $S = 56$



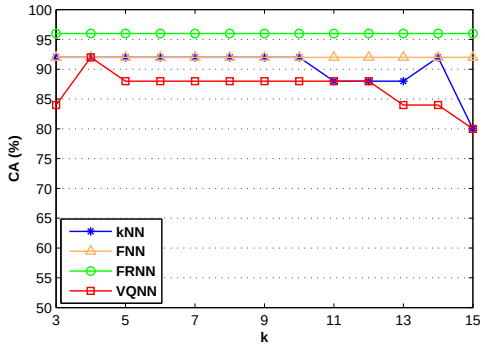
(b) A_z for MIAS (manual annotation), $S = 56$



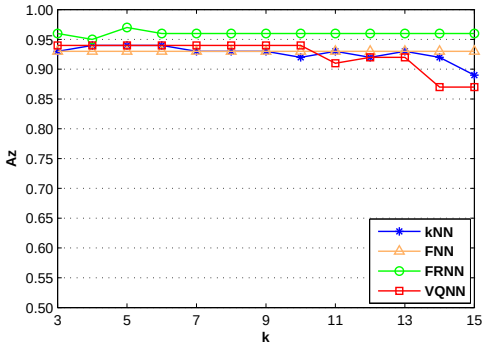
(c) CA for MIAS (automatic detection), $S = 6$



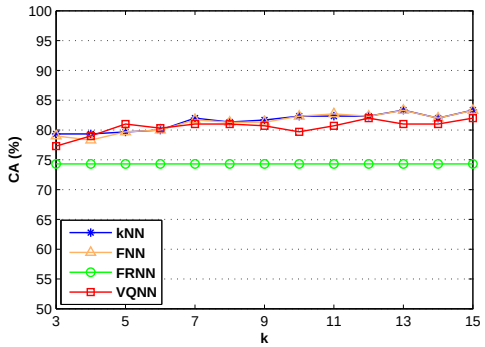
(d) A_z MIAS (automatic detection), $S = 6$



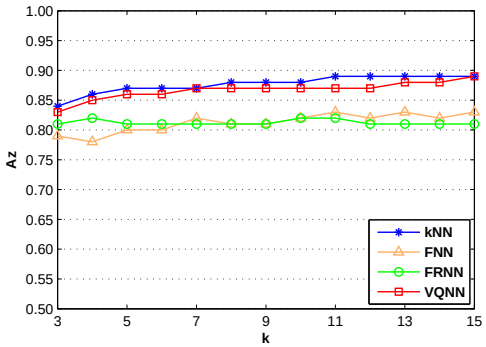
(e) CA for Digital, $S = 16$



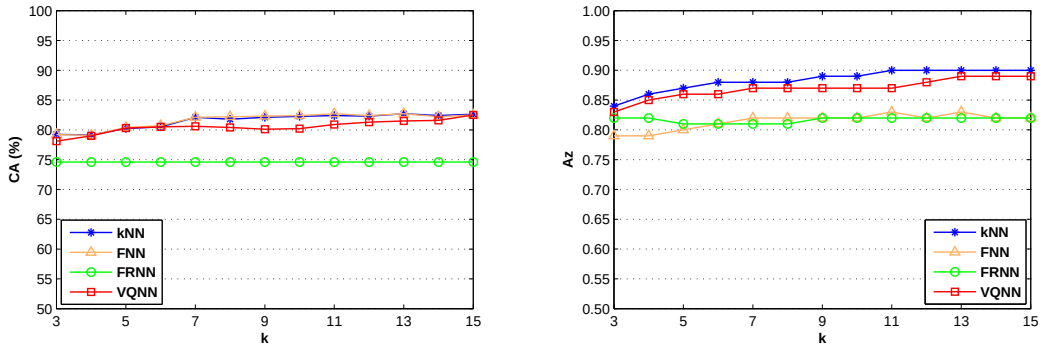
(f) A_z for Digital, $S = 16$



(g) CA for DDSM (LOOCV), $S = 58$



(h) A_z for DDSM (LOOCV), $S = 58$



(i) CA for DDSM (10-FCV), $S = 58$ (j) A_z for DDSM (10-FCV), $S = 58$

Figure 7.11: The classification results of the four classifiers with respect to different k values for the three datasets. Left column: CA with respect to different k values; right column: A_z with respect to different k values. Note that some curves for different classifiers are overlapped.

experiments by combining both features compared to using either of them). The former represents the number of independent objects resulted from dilating each individual microcalcification at a certain scale. The latter describes the extent of the interior connectivity within these objects. The topological features extracted over a range of scales were aggregated to generate a multiscale topological feature vector, which was used in the classification procedure.

The validity of our method has been evaluated using three datasets: MIAS, DDSM and Digital. Four k-Nearest Neighbours based algorithms (kNN, FNN, FRNN and VQNN) were used for the classification. Good classification results have been obtained for all the datasets. By using a sequence of feature subsets corresponding to an increasing scale range, the obtained best classification accuracy was 95% for MIAS with manual annotations, 95% for MIAS with automatic detections, 96% for Digital, 84.3% for DDSM using leave-one-out cross validation, and $83.8\% \pm 5.7$ for DDSM based on 10-fold cross validation; and the largest area under the ROC curve was 0.96, 0.96, 0.97, 0.89 and 0.90 ± 0.06, respectively. The Digital dataset provided the best results, which might be due to the more accurate detection of microcalcifications using digital mammography. As stated in (Oliver *et al.*, 2012), the detection approach indicates the best performance when using the Digital dataset, and therefore more accurate detection results of microcalcifications can be provided for our classification task. For the MIAS dataset, the second best classification results were indicated. When using the manual segmentation of microcalcifications, the obtained results were generally better than those of using the automatic detection of microcalcifications over most scale ranges, probably due to the more precise identification of visible microcalcifications by the expert, but the same best performance was achieved. This indicates that the proposed method

seems robust with respect to variations between manual and automatic segmentations of microcalcifications. For the DDSM dataset, very similar results were shown when using the leave-one-out and 10-fold cross-validation methods, showing a decreased performance in the results when compared to the other datasets. It might be partially explained by the fact that the detection approach performs worst for the DDSM database among the four datasets (Oliver *et al.*, 2012). Moreover, the DDSM dataset used in our experiments contains 300 cases, which might induce a larger variability than the small datasets (especially as the DDSM dataset was generated using different digitisers). However, the obtained classification results are still comparable or even better than the related work reviewed in (Cheng *et al.*, 2003), where most publications used smaller databases than ours.

We compared our proposed method with some related publications. Table 7.3 shows a summary of the comparison. Overall, the obtained classification results are comparable to the various approaches. Note that those approaches in Table 7.3 use different images taken from different databases, and therefore it is a qualitative comparison. The drawbacks of some approaches have to be pointed out though they have indicated superior results. In (Shen *et al.*, 1994), the 100% CA was obtained by classifying 143 individual microcalcifications from 18 biopsy proven cases based on a leave-one-microcalcification-out approach, which is different from the goal of our classification of microcalcification clusters. In (Ma *et al.*, 2010), the classification of microcalcification clusters was based on the maximum feature value obtained by a selected microcalcification rather than the whole cluster (and some manual aspects were involved in the extraction process). In (Ren *et al.*, 2011), the high classification performance was obtained by introducing an optimised decision making step which was performed afterwards through statistical analysis of the classifier outputs to achieve the minimum cost of error classification. One limitation of this method is that it cannot provide a reliable classification for the case where the cluster is structureless or few microcalcifications are segmented within the cluster. An extreme is that if only a single microcalcification is detected from the cluster by the automatic detection approach, it will fail to discriminate malignant from benign based on the topology. Another inherent limitation of this method is that its performance depends on the sensitivity and specificity of the microcalcification detection. False negatives and false positives in the detection results may change the global topology/connectivity of microcalcification clusters. However, the experimental results show the validity of our method when combined with automatic microcalcification detection.

As discussed above, some highlights of this work should be noted. For the MIAS dataset, the same classification methodology was applied based on both manually annotated and automatically detected microcalcifications, and similar classifica-

Table 7.3: Comparison of our results with those achieved by related work.

Method	Database	Case	Feature	Classifier	Result
Shen *et al.* (1994)	unknown	18	shape	kNN	$CA = 100\%$
Ma *et al.* (2010)	DDSM	183	shape	threshold	$A_z = 0.96$
Chan *et al.* (1998)	unknown	145	morphology	LDC	$A_z = 0.79$
Dhawan *et al.* (1996)	unknown	191	texture&cluster	ANN	$A_z = 0.86$
Papadopoulos *et al.* (2005)	MIAS	25	cluster	SVM	$A_z = 0.81$
Chan *et al.* (1997)	unknown	54	texture	ANN	$A_z = 0.88$
Soltanian-Zadeh *et al.* (2004)	Nijmegen	103	multiwavelet	kNN	$A_z = 0.89$
Betal *et al.* (1997)	Liverpool	38	shape/cluster	kNN	$A_z = 0.79, A_z = 0.84$
Rana *et al.* (2007)	University of Chicago Hospitals	49	morphology	ANN	$A_z = 0.80$
Wei *et al.* (2009)	University of Chicago	104	cluster&morphology	Ada-/Cas-SVM	$A_z = 0.81, A_z = 0.82$
Shao *et al.* (2011)	Sun Yat-sen University	109	pattern form factor	threshold	$A_z = 0.74$
Ren *et al.* (2011)	DDSM	150	varied features	improved ANN	$A_z = 0.98$
Ren (2012)	DDSM	150	varied features	improved ANN/SVM	$A_z = 0.93, A_z = 0.94$
Ours	MIAS I (manual annotation)	20	topology	FNN	$CA = 95\%, A_z = 0.96$
Ours	MIAS I (automatic detection)	20	topology	FNN	$CA = 95\%, A_z = 0.96$
Ours	Digital	25	topology	FRNN	$CA = 96\%, A_z = 0.97$
Ours	DDSM (leave-one-out CV)	300	topology	kNN	$CA = 84.3\%, A_z = 0.89$
Ours	DDSM (10-fold CV)	300	topology	kNN	$CA = 83.8 \pm 5.7\%, A_z = 0.90 \pm 0.06$

tion performance was obtained. This indicates the robustness of this method to detection errors. For the DDSM dataset, we used a larger set of cases than related publications and achieved good results. In addition to the digitised databases, we evaluated our method using a full-field digital database and obtained improved classification results. This demonstrates the capability of our method in dealing with two categories of mammograms, which allows it to be applied in both film and digital mammography. On the other hand, we used the CAD detection results directly instead of manual segmentation results for all the datasets. High classification accuracies and good ROC analysis results were obtained compared to state of the art approaches. This indicates its potential application in CAD systems.

As future work, a range of other properties of microcalcification graphs can be investigated to generate a more sophisticated representation of microcalcification clusters. Other features such as shape and texture of individual microcalcifications and the whole cluster could be incorporated to build a complete microcalcification analysis framework. Weighted graphs may be applied to involve the spatial distance between two microcalcifications. The definition of a similarity measure between graphs can be investigated in order to realise classification using the graph based representation directly without generating feature vectors. On the other hand, alternative classifiers (e.g. random forests, ANN, and SVM) could also be investigated. In addition, we will extend the evaluation using a larger collection of digital mammograms.

Chapter 8

Discussions

In Chapter 6, an evaluation of the validity of all the proposed methods for mammographic risk assessment has been conducted. In this chapter, we present a comprehensive analysis on the obtained results. Firstly, a comparison is made among our proposed methods where classification performance of individual methods is compared. Subsequently, the performance of our proposed methods and that of related publications described in Chapter 2 are compared. The reasons for the difference in the performance of the various approaches are discussed. In addition, possible directions for further work are discussed.

8.1 Comparison of the Proposed Methods

In the previous chapters, we have proposed two breast density segmentation methods, five local feature based strategies for breast tissue appearance modelling, and one multiscale blob based approach for representing parenchymal patterns in mammograms. Figure 8.1 shows the resulting segmentation/representation of four example mammograms and confusion matrices for the BIRADS breast density classification by using different approaches. On the left side of Figure 8.1, the top row is the original mammograms without any pre-processing; the second row shows the breast density segmentation results based on the modified fuzzy c-means (MFCM) algorithm where the breast is segmented into six different densities (different colours represent different densities and from dark blue to dark red corresponds to an increasing density); the third row indicates the resulting *LBP* label maps when using 5×5 local windows (different colours show different local binary patterns); the fourth row corresponds to the resulting *LGA* label maps when using 5×5 local windows and a greylevel resolution of 16 (different colours show different local greylevel appearances); the fifth row is the resulting *BIF* label maps when $\sigma_{base} = 1$ and $\varepsilon = 0$ (different colours represent different configurations

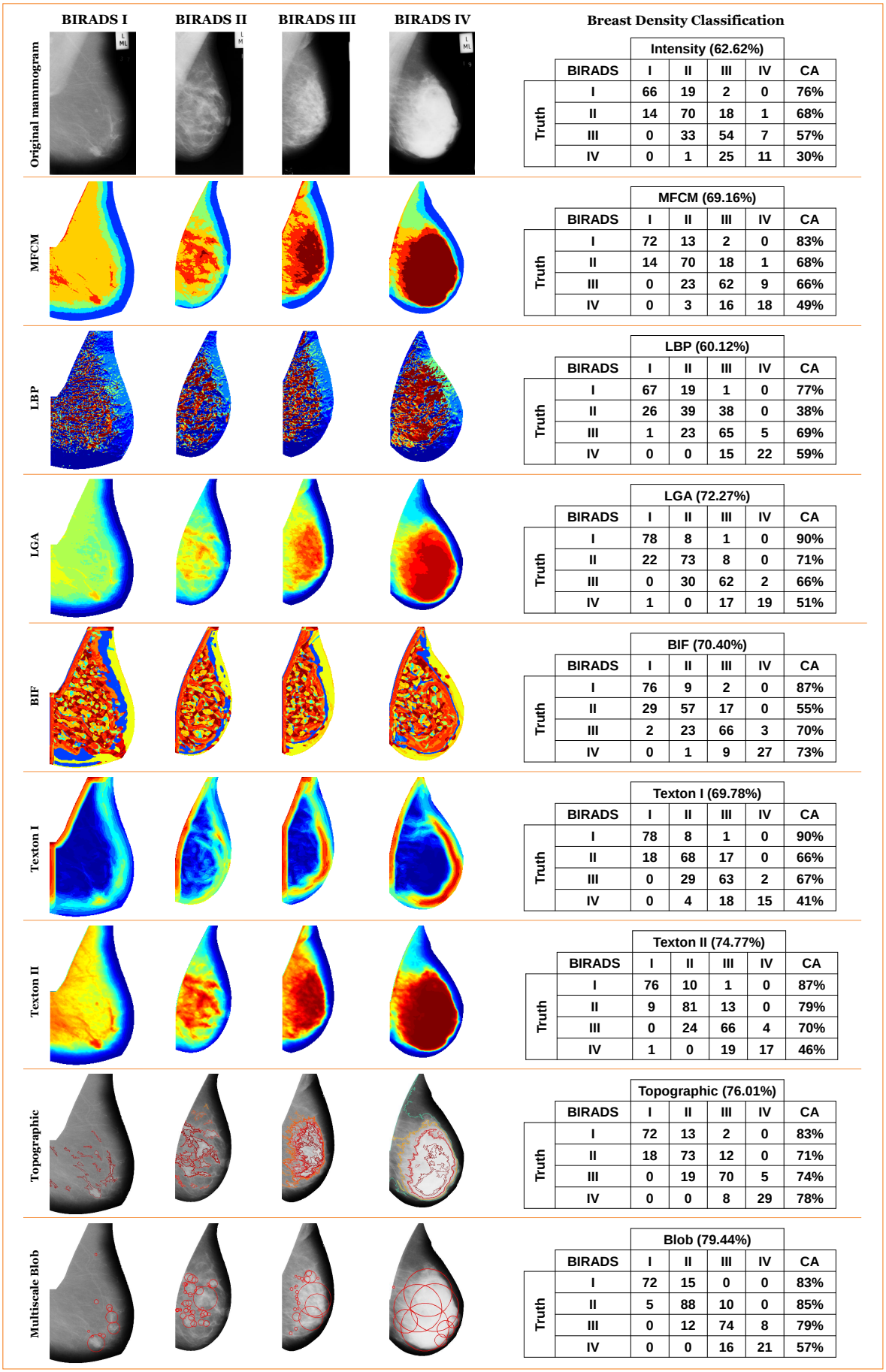


Figure 8.1: A comparison among the proposed methods.

of a stack of *BIFs* over four scales); the sixth and seventh rows show the resulting label maps corresponding to 160 textons for Texton I and Texton II (different colours represent different textons); the eighth row shows the breast density segmentation results based on the topographic approach where dense tissue regions are detected and their boundaries are delineated using colour curves (different colours correspond to different densities); and the bottom row indicates the resulting blobs detected from different scales. On the right side of Figure 8.1, the breast density classification results for the full MIAS database are given, where the top confusion matrix is obtained by using the intensity histograms of the original mammograms as feature vectors for breast density classification, and the remaining confusion matrices correspond to the proposed methods in the same row, taken from Tables 6.3(b), 6.8(a)-(e), 6.5(a), and 6.13(b).

As indicated in Figure 8.1, the best classification performance was achieved by the multiscale blob based approach, providing an overall classification accuracy (CA) of 79.44% for the four BIRADS density categories. The second-best classification result was obtained by the topography based breast density segmentation, providing an overall CA of 76.01%. This was followed by Texton II and a CA of 74.77% was produced, which was the best performance among all the five local feature based strategies. Following Texton II, LGA also indicated a good classification result and a CA of 72.27% was obtained. After that, MFCM, BIF, and Texton I produced similar results around 70%, and the resulting CA was 69.16%, 70.40%, and 69.78%, respectively. When using the original intensity histograms, only a 62.62% CA was achieved, which was much lower than those of the approaches above. However, the worst performance was shown by LBP, which only provided an overall CA of 60.12%.

The reasons why the blob based approach and the topography based approach indicated superior performance to the other approaches may be that they both focused on dense tissue in the breast (which plays an important role in breast density classification) and detected region-wise dense tissue as multiscale blobs or prominent/independent *shapes* (connected components) in the topographic map. The resulting dense tissue regions (blobs or *shapes*) over a range of scales can be well linked to the definition of the BIRADS density classification. In particular, as shown in the last two rows of Figure 8.1, for BIRADS I, the breast is almost entirely fatty, and only few small scale blobs or *shapes* were detected; for BIRADS II, there is some scattered fibroglandular tissue in the breast, and a few scattered blobs or *shapes* with small or medium scales were detected around the fibroglandular tissue; for BIRADS III, the breast is heterogeneously dense, and relatively larger scale blobs or *shapes* were present in the resulting regions; while for BIRADS IV, the breast is extremely dense, and large scale blobs or *shapes* were detected, covering

the main area of the breast.

Texton II and LGA showed the second-best performance among the remaining approaches, which may be explained by the fact that they both concentrated more on the density aspect to depict breast tissue appearance when compared to the other three local feature based approaches. As shown in the fourth and seventh rows of Figure 8.1, their resulting label maps were closely associated with the density based segmentation results shown in the second row of Figure 8.1. As two related methods in the sense that they both aimed to model local texture information based on the joint greylevel distribution within the local neighbourhoods, the slightly worse performance of LGA than Texton II may be due to the fact that in contrast to Texton II using k-means clustering to learn the cluster centres of the local neighbourhoods and as such reduce the dimensionality of feature space, there was no initial learning step in LGA to learn the main components of all possible greylevel appearances within the local neighbourhoods, and thus the classification was performed on the full feature space which was much sparser and might introduce more noise than Texton II.

For BIF and Texton I, they focused more on the structural aspect to describe breast tissue appearance. Filter banks consisting of a set of local window filters were employed to detect various structures of different orientations/scales within the local neighbourhoods, such as edges, bars, blobs, flat patterns, and gradient patterns, etc. Although greylevel information was included to some extent in the process of local window filtering, structural information played a dominant role in representing local tissue appearance. This resulted in entirely unrelated label maps to the density based segmentation as shown in the fifth and sixth rows of Figure 8.1. Some of these geometric structures in breast tissue are relatively less significant for evaluating the overall breast density and thus BIF and Texton I indicated the intermediate results for breast density classification.

MFCM was simply based on two implicit assumptions: one was that the same greylevel value indicates the same tissue density over different mammograms and thus pixels with similar greylevel values belong to the same density type; the other was that mammograms belonging to the same BIRADS class have similar tissue compositions (relative proportions of tissue with different densities). However, there is a variability in intensity values with respect to the same tissue density among different mammograms, due to the variance in the mammogram acquisition process. Moreover, mammograms in different BIRADS classes may have approximately similar amounts of tissue regarding some density types (e.g. a band of fatty tissue along the breast boundary). This may induce confusion among different BIRADS classes especially for the neighbouring classes. Thus, MFCM obtained worse performance by only segmenting the breast into six density sub-regions than

Texton II and LGA (which provided a more detailed description for breast tissue density by incorporating spatial information into greylevel information).
Similarly, using the original intensity histograms directly for breast density classification was also based on an implicit assumption that mammograms belonging to the same BIRADS class have similar greylevel distributions. However, mammograms in different BIRADS classes may have relatively similar intensity histograms due to the statistical nature of histograms without taking into account spatial information. This is a primary cause of the misclassification between the neighbouring classes and may explain the poor performance.
LBP mainly focused on the difference between the neighbouring pixels and the centre pixel in the local neighbourhoods and transformed every local neighbourhood into a binary pattern, which excluded greylevel information from the resulting breast tissue models. However, greylevel information plays an important role in describing breast tissue density, and therefore it indicated the worst performance among all these approaches.
Furthermore, we investigated the performance of combining some approaches above for breast density classification using the MIAS database. We implemented the combination of different approaches in two ways. The first was combining the outputs of the five individual classifiers built by the five local feature based strategies (i.e. LBP, LGA, BIF, Texton I, and Texton II) using a weighted average rule (see Section 6.3.2); and the second was concatenating the feature vectors derived from the MFCM based approach and the Texton II approach to generate new feature vectors (see Section 6.3.3). The improved results were indicated by using the combination of classifiers or features in different approaches compared to using these approaches individually (see Tables 6.8 and 6.10).
In addition, we applied the topography based method for breast density segmentation and the multiscale blob based method for parenchymal pattern representation to the DDSM database, in order to evaluate the robustness of these two methods based on a different and larger dataset (since they showed better performance than the other proposed methods for the MIAS database) (see Sections 6.3.1.2 and 6.3.4). When using features derived from single-view (MLO or CC) mammograms for the BIRADS density classification, the achieved CAs for the DDSM database were slightly lower than those for the MIAS database but still indicated the robustness/independence of these two proposed methods on a large dataset; moreover, when combining features derived from two-view (MLO/CC) mammograms for classification using the topographic breast density segmentation method, the CA for the DDSM database was much improved and even higher than the best CA for the MIAS database (see Tables 6.5, 6.13 and 6.14).
Table 8.1 shows a summary of the results obtained by the proposed methods for

Table 8.1: A summary of the classification results for the MIAS and DDSM databases, obtained by using our proposed methods and the leave-one-woman-out cross validation scheme. B-I, B-II, B-III, and B-IV represent the four BIRADS classes (BIRADS I-IV); Low and High represent the low density and high density classes; Overall_4 and Overall_2 stand for the overall CAs for the four-class and two-class density classification, respectively; Comb. I indicates the combination of LBP, LGA, BIF, Texton I, and Texton II; Comb. II indicates the combination of MFCM and Texton II; $\langle\cdot\rangle_M$, $\langle\cdot\rangle_C$, and $\langle\cdot\rangle_{MC}$ mean performing the approach on MLO, CC, and both MLO and CC mammograms, respectively. The results for Comb. I_M and Comb. II_M are taken from Tables 6.8(f) and 6.11(j), respectively. The results of the topographic approach for the DDSM database are obtained from Tables 6.5 (b)-(d).

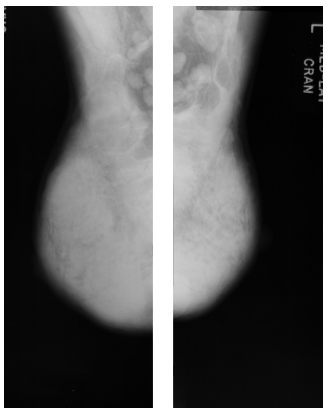
		Classification Accuracy (%)							
Database	Method	B-I	B-II	B-III	B-IV	Overall_4	Low	High	Overall_2
MIAS	MFCM_M	83	68	66	49	69.16	89	80	85.36
	LBP_M	77	38	69	59	60.12	79	82	80.37
	LGA_M	90	71	66	51	72.27	95	76	87.54
	BIF_M	87	55	70	73	70.40	90	80	85.98
	Texton I_M	90	66	67	41	69.78	91	75	84.11
	Texton II_M	87	79	70	46	74.77	93	81	87.85
	Topographic_M	83	71	74	**78**	76.01	93	86	89.72
	Blob_M	83	**85**	**79**	57	79.44	95	91	93.15
	Comb. I_M	**94**	73	76	59	77.88	95	84	90.34
	Comb. II_M	85	74	65	54	71.96	90	79	85.36
DDSM	Blob_M	**78**	**86**	71	64	76.90	94	85	89.77
	Topographic_M	75	79	**73**	**69**	75.21	92	89	90.25
	Topographic_C	68	83	71	66	74.61	94	85	89.65
	Topographic_{MC}	88	85	78	72	81.23	95	88	91.70

the MIAS and DDSM databases, where the left part (split by the double lines) indicates the CAs for each BIRADS class and the overall CAs for the four-class BIRADS density classification, while the right part indicates the CAs for the low density (BIRADS I/II) and high density (BIRADS III/IV) classes as well as the overall CAs for the two-class (low/high) density classification.

For the MIAS database, the best overall CA for the BIRADS density classification was 79.44% achieved by the multiscale blob based method, and this increased to 93.15% for the low/high density classification. The combination of the five local feature based strategies also showed good classification results with an overall CA of 77.88% for the BIRADS density classification and an overall CA of 90.34% for the low/high density classification. This was closely followed by the topographic approach, which obtained 76.01% and 89.72% overall CAs for the BIRADS and low/high density classification, respectively. In addition, Texton II and LGA indicated promising results, providing over 72% overall CAs for the BIRADS density

classification and over 87% overall CAs for the low/high density classification. When concentrating on each BIRADS class, the best CA for BIRADS I was 94% obtained by Comb. I, and the average CA over the ten proposed methods (ranging from MFCM to Comb. II in Table 8.1) was $85.90 \pm 4.79\%$. For BIRADS II, the multiscale blob based method produced the best CA of 85%, and the average CA over all the proposed methods was $68.00 \pm 13.17\%$ (this large standard deviation was caused by the poor performance of LBP and BIF). For BIRADS III, the best CA was 79% also produced by the blob based method, and the average CA was $70.20 \pm 4.71\%$. For BIRADS IV, the best CA was 78% obtained by the topographic method, and the average CA over the proposed methods was $56.70 \pm 11.50\%$. All the proposed methods except Blob indicated the best performance for BIRADS I, which tend to imply that mammograms in BIRADS I are relatively easy to distinguish from those in the other three BIRADS classes due to lower variance in mammographic tissue appearnce. In contrast to BIRADS I, mammograms in BIRADS II and III showed higher variance and therefore increased the difficulty in modelling mammographic tissue appearance for these two classes. This may explain the worse performance for BIRADS II and III compared to BIRADS I. In addition, most of the proposed methods (excluding LBP, BIF, and Topographic) performed worst for BIRADS IV among the four BIRADS classes, where a large number of mammograms were misclassified into BIRADS III. This might be partly due to the unbalanced distribution of these two classes in the dataset. On the other hand, some BIRADS IV mammograms indicated structureless and low contrast tissue nearly all over the entire breast region, and moreover showed low intensity values (two examples are given in Figure 8.2), which resulted in failing to provide enough information for modelling breast tissue characteristics from a computer vision point of view. This might be another reason for the poor performance of most of the above methods for BIRADS IV. When considering the low/high density classification, all the proposed methods except LBP obtained better results for the low density class than those for the high density class, and the average CAs were $91.00 \pm 4.78\%$ and $81.40 \pm 4.72\%$, respectively.

For the DDSM database, the best overall CA for the BIRADS density classification was 81.23%, achieved by Topographic$_{MC}$ where the features derived from the MLO and CC mammograms of the same breast were combined for classification, and this also achieved the best result for the low/high density classification, providing an overall CA of 91.70%. For the topographic method, when using the features derived from the MLO and CC mammograms individually for classification, better results were obtained for the MLO mammograms with an overall CA of 75.21% for the BIRADS density classification and an overall CA of 90.25% for the low/high density classification. For the MLO mammograms, Blob$_M$ produced a better



(a) mdb062rs (b) mdb061ls

Figure 8.2: Two example misclassified BIRADS IV mammograms from MIAS.

overall CA of 76.90% for the BIRADS density classification, while Topographic$_M$ provided a better overall CA of 90.25% for the low/high density classification. When considering each BIRADS class, we focused on the results for the MLO mammograms. Blob$_M$ indicated better CAs of 78% and 86% for BIRADS I and II, while Topographic$_M$ achieved better CAs of 73% and 69% for BIRADS III and IV. The average CAs of these two methods for the four BIRADS classes were 76.50±2.12%, 82.50±4.95%, 72.00±1.41%, and 66.50±3.54%, respectively. Thus, in contrast to their obtained results for the MIAS database, the best performance was shown by BIRADS II instead of BIRADS I, and BIRADS I indicated the second-best performance. Similarly, BIRADS III and IV showed relatively lower performance, and BIRADS IV indicated the lowest performance. The difference in classification performance for the individual BIRADS classes between the MIAS and DDSM databases might be partly due to the different distribution of the four BIRADS classes in the two databases. In addition, the inter/intra-class variance in mammographic tissue appearance is different between the two databases, which may also result in the different classification performance with respect to the same BIRADS class. When analysing the low/high density classification, better results were obtained for the low density class compared to the high density class, and the average CAs of Blob$_M$ and Topographic$_M$ were $93.00 \pm 1.41\%$ and $87.00 \pm 2.83\%$ for the low density class and the high density class, respectively.

When comparing the obtained results for the MIAS and DDSM databases, we focused only on the results of Blob$_M$ and Topographic$_M$ in order to make a fair comparison. Slightly worse results were obtained for the DDSM database compared to those for the MIAS database. As mentioned earlier, this might be due to the fact that the DDSM database contains a larger number of mammograms than the MIAS database and as such it indicates larger intra-class variance which may affect the final classification performance.

8.2 Comparing with Related Publications

In the above section, we have compared the experimental results obtained using our proposed methods and discussed the reasons for the different performance of these methods. In this section, we compare the obtained results with those obtained by related publications in the literature. A number of publications were selected from the various works reviewed in Chapter 2 for this comparison, which we believe are the most representative/advanced methods for mammographic risk assessment in the state of the art. A complete comparison of the proposed methods with related publications in the literature can be found in Table 8.2.

Bovis & Singh (2002) evaluated their developed method using 377 MLO mammograms from the DDSM database. When using the eleven ANN classifiers individually, the best CA was 58.3%. When combining the posterior probabilities of the eleven classifiers, an increased recognition rate of 71.4% was obtained using the product rule. Petroudi *et al.* (2003) selected 132 mammograms from the Oxford Database for the evaluation, where 44 mammograms were used as the training set and the remaining 88 mammograms were used as the testing set. An overall CA of 76% was obtained based on the BIRADS ground truth provided by one breast radiologist and two image analysts. Oliver *et al.* (2005a) used a set of 300 MLO mammograms (50 for BIRADS I and IV, 100 for BIRADS II and III) taken from the DDSM database for evaluating the proposed method. A leave-one-image-out cross-validation methodology was applied, and the overall CA was 40.3% for kNN, and 43.3% for ID3, respectively. This increased to 47% when combining the two classifiers by averaging their outputs. In the study of Oliver *et al.* (2006), the evaluation was based on 322 MLO mammograms in the MIAS database using a leave-one-woman-out methodology. The obtained CAs for five different strategies were 62%, 79%, 82%, 81%, and 80%, respectively. The ground truth for the BIRADS density classification was from only one expert.

Castella *et al.* (2007) proposed a semi-automatic method and tested their method on a non-public database collected at the Clinique des Grangettes, Geneva, Switzerland, consisting of 352 digital mammograms from 88 women (both MLO and CC view mammograms were taken per breast). The first step of this method was the manual selection of four ROIs per mammogram by radiologists and thus 1408 ROIs were obtained. Each ROI was then classified as one BIRADS density class using a leave-one-ROI-out methodology. The BIRADS density label for each mammogram was determined by averaging the individual classification for the four ROIs. Fisher linear discriminant was employed to select the optimal feature subspace and improve classification performance. The best overall CAs obtained using LDA, Naive Bayesian, and Mahalanobis Bayesian were 80%, 55%, and 69%, respectively.

Table 8.2: A complete comparison of the proposed methods with related publications in the literature, including the aspects of feature, classifier, risk assessment scheme, experimental data, evaluation methodology, and overall classification accuracy. LOO[1], LOO[2], LOO[3], and n-F represent leave-one-image-out, leave-one-woman-out, leave-one-ROI-out, and n-fold cross-validation, respectively.

Method	Feature	Classifier	Scheme	Data	Evaluation	CA
Bovis & Singh (2002)	texture+statistical+shape+fractal	1 ANN, 11 ANN	BIRADS	377 MLO DDSM	10-F	58.3%, 71.4%
Petroudi *et al.* (2003)	texton histogram	Nearest Neighbour	BIRADS	132 MLO/CC Oxford	44/88	76%
Oliver *et al.* (2005a)	morphological+texture	kNN, ID3, kNN+ID3	BIRADS	300 MLO DDSM	LOO[1]	40.3%, 43.3%, 47%
Oliver *et al.* (2006)	morphological+texture	Bayesian (kNN+C4.5)	BIRADS	322 MLO MIAS	LOO[2]	62%, 79%, 82%, 81%, 80%
Castella *et al.* (2007)	statistical	LDA, Bayesian	BIRADS	352 MLO/CC Swizerland	LOO[3]	80%, 55%, 69%
Oliver *et al.* (2008)	morphological+texture	SFS+kNN, C4.5, Bayesian	BIRADS	322 MLO MIAS	LOO[2]	77%, 72%, 86%
Oliver *et al.* (2008)	morphological+texture	SFS+kNN, C4.5, Bayesian	BIRADS	831 MLO DDSM	LOO[2]	70%, 72%, 77%
Mac Parthaláin *et al.* (2010)	morphological+texture	DMTRS+FRNN	BIRADS	322 MLO MIAS	10-F	91.4%
Mac Parthaláin *et al.* (2010)	morphological+texture	DMTRS+FRNN	BIRADS	831 MLO DDSM	10-F	89.24%
Subashini *et al.* (2010)	statistical	SVM	Three Categories[1]	43 MLO MIAS	3-F	95.44%
He *et al.* (2010)	Tabár's model based segmentation	kNN	Tabár, BIRADS	320 MLO MIAS	LOO[2]	78%, 75%
Bueno *et al.* (2011)	statistical	PCA+kNN	BIRADS	322 MLO MIAS	H-O+R-S	84%
Bueno *et al.* (2011)	statistical	PCA+LBN+SVM	BIRADS	322 MLO MIAS	10-F	75%
Tzikopoulos *et al.* (2011)	fractal dimension+statistical	CARTs, kNN, SVM	Three Categories[1]	322 MLO MIAS	LOO[2]	65.84%, 76.40%, 77.02%
He *et al.* (2011)	Tabár's model based segmentation	Nearest Neighbour	Tabár, BIRADS	320 MLO MIAS	LOO[1]	53%, 70%
He *et al.* (2012)	density based segmentation	kNN	Tabár, BIRADS	321 MLO MIAS	LOO[2]	85%, 78%
MFCM (Sec. 3.3)	intensity	kNN	BIRADS	321 MLO MIAS	LOO[2]	69.16%
LBP (Sec. 4.3)	local binary pattern			321 MLO MIAS	LOO[2]	60.12%
LGA (Sec. 4.3)	local greylevel appearance			321 MLO MIAS	LOO[2]	72.27%
BIF (Sec. 4.3)	local geometric structure			321 MLO MIAS	LOO[2]	70.40%
Texton I (Sec. 4.3)	local texture			321 MLO MIAS	LOO[2]	69.78%
Texton II (Sec. 4.3)	raw image patch			321 MLO MIAS	LOO[2]	74.77%
Topographic (Sec. 3.4)	prominent/independent *shape*			321 MLO MIAS	LOO[2]	76.01%
Blob (Sec. 5.3)	blob-like image structure			321 MLO MIAS	LOO[2]	79.44%
Comb. I (Sec. 6.3.2)	local features			321 MLO MIAS	LOO[2]	77.88%
Comb. II (Sec. 6.3.3)	intensity+raw image patch			321 MLO MIAS	2-F	71.96%
Blob (Sec. 5.3)	blob-like image structure			831 MLO DDSM	LOO[2]/10-F	76.90%/76.15 $\pm$ 0.23%
Topographic (Sec. 3.4)	prominent/independent *shape*			831 MLO DDSM	LOO[2]/10-F	75.21%/75.09 $\pm$ 0.62%
Topographic (Sec. 3.4)	prominent/independent *shape*			831 CC DDSM	LOO[2]/10-F	74.61%/74.34 $\pm$ 0.72%
Topographic (Sec. 3.4)	prominent/independent *shape*			1662 MLO/CC DDSM	LOO[2]/10-F	81.23%/80.88 $\pm$ 0.44%

Oliver *et al.* (2008) applied their previous method to 831 MLO mammograms from the DDSM database in addition to the full MIAS database. Three classifiers, kNN, C4.5 decision tree, and the Bayesian combination of kNN and C4.5, were used for the BIRADS density classification, and the sequential forward selection (SFS) algorithm was applied to select the optimal feature subset containing the most discriminant features. For the MIAS database, the best performance of the individual classifiers was 77%, 72%, and 86%; and for the DDSM database, the performance obtained by the three classifiers was 70%, 72%, and 77%, respectively. These classification results were obtained according to the consensus ground truth of three experts (the same ground truth as used in this thesis).

Mac Parthaláin *et al.* (2010) employed a number of sophisticated kNN classifiers (FNN, FRNN, FRNN-O, and VQNN) to process the resulting feature data of Oliver *et al.* (2008), incorporating fuzzy and fuzzy-rough approaches into the classic kNN classification. A 10-fold cross-validation methodology was used for evaluation. Two feature selection methods (FRFS and DMTRS) were applied and the large improvement of CA was achieved for all classifiers using the reduced feature space. The FRNN classifier combined with the DMTRS approach indicated the best performance. The resulting CAs were up to 91.4% and 89.24% for the MIAS and DDSM databases, respectively.

Subashini *et al.* (2010) selected 43 normal mammograms from the MIAS database and classified them into three density categories using a SVM classifier. The performance of different combinations of fourteen extracted features was tested to select the optimal subset of nine features. An overall CA of 95.44% was achieved by using a three-fold cross-validation method.

Bueno *et al.* (2011) tested the proposed method on the full MIAS database. A two-layer tree classifier was designed for the four-class BIRADS density classification, where the first layer was intended to partition BIRADS I/II and BIRADS III/IV, and the second layer was intended to further discriminate BIRADS I from II and BIRADS III from IV. Three classifiers were incorporated into the hierarchical classification procedure, and the classifiers were trained and tested in two ways. Specifically, the kNN classifier was first used for both layers and the combination of the hold-out (H-O) and re-substitution (R-S) methods was employed for training and testing (the data was first randomly split into two groups containing the same number of samples and one of them was randomly selected as the training set, then the test was performed on the whole dataset including the training data), which produced an overall CA of 84%. On the other hand, linear Bayes normal (LBN) was used for the first layer and SVM was used for the second layer, and 10-fold cross-validation was employed for training and testing. This provided a decreased CA of 75% compared to that obtained earlier. In addition, PCA was applied

to reduce the dimensionality of feature space. Different tests were performed by varying the number of the resulting components by PCA, in order to select the optimal number of components and minimise the classification error. The BIRADS classification ground truth was obtained from expert clinicians in a local hospital. Tzikopoulos *et al.* (2011) also evaluated their developed method using the full MIAS database. Breast density classification was performed according to the three density categories provided in the MIAS database. When using a leave-one-woman-out cross-validation methodology for the evaluation, the best CAs obtained by CARTs, kNN, and SVM were 65.84%, 76.40%, and 77.02%, respectively.

In addition, He *et al.* (2010, 2011, 2012) used the MIAS database to evaluate the capability of the resulting segmentation for mammographic risk assessment. Mammographic images were classified according to both Tabár and BIRADS categories. When using the texture signature based segmentation method (He *et al.*, 2010), the obtained CAs were 78% and 75% for the Tabár and BIRADS classification, respectively. When using the segmentation method based on geometric moments (He *et al.*, 2011) and a leave-one-image-out cross-validation methodology, 53% and 70% correct classification were obtained for the Tabár and BIRADS classification, respectively. Significantly improved results were achieved in their latest work (He *et al.*, 2012), and the CAs were increased to 85% and 78% for Tabár and BIRADS, respectively. For the BIRADS density classification, the ground truth used in He *et al.* (2010) was from one expert, while the ground truth used in He *et al.* (2011, 2012) was the concensus classfication of three experts (the same ground truth as used in this thesis).

As shown in Table 8.2, our proposed methods achieved good results for both MIAS and DDSM databases compared to related publications. It is not always possible to make a direct comparison with related work in the literature, as the difference in the database and/or ground truth can affect the classification results. Among these related publications in Table 8.2, Oliver *et al.* (2005a, 2008) and He *et al.* (2011, 2012) conducted the evaluation based on the same database and the same BIRADS classification ground truth as those used in this thesis, which allowed a direct comparison. While for the remaining publications, note that it is only a qualitative comparison due to the use of different databases and/or ground truth for the evaluation. It is also noteworthy that some of the publications above employed advanced classifiers or a combination of multiple classifiers for the classification, and applied dimensionality reduction techniques such as SFS, PCA, LDA, and DMTRS to the classification process to achieve superior classification performance. This could raise the risk of overfitting the data and result in overfitted classification models. In particular, the SFS algorithm is widely known as a wrapper feature selection method, which uses classification performance to evaluate

features and selects a feature subset well fit for the given classifier. On the other hand, some approaches performed feature selection in a way of brute force testing (e.g. (Castella *et al.*, 2007; Subashini *et al.*, 2010; Bueno *et al.*, 2011)), where the performance of every possible combination of the attributes in feature vectors was tested in order to select an optimal subset of features that can provide the best classification performance. In addition, in most of the publications (if not all) involving feature selection, it seems that the feature selection process was performed on the complete dataset. This would clearly induce overfitting in the sense that the classification model is built and validated using the same data. When focusing on the aspect of training and testing, the implementation of some approaches was not on the basis of completely separating the training data from the testing data. However, any overlap between the training data and the testing data is clearly not feasible in clinical practice. Specifically, Bueno *et al.* (2011) used the combined method of hold-out and re-substitution for training and testing, which involved all the training samples into the test set. Castella *et al.* (2007) used the leave-one-ROI-out methodology and integrated the individual ROI classification results into an overall density rating for each mammogram. This could induce bias in the classification of mammograms as the ROIs selected from the same mammogram might contain similar breast tissue. He *et al.* (2011) used the leave-one-image-out methodology when evaluating their method on the MIAS database, which could also cause bias due to the fact that left and right mammograms of the same woman tend to have similar tissue features. In contrast, we only used the kNN classifier for the classification and did not involve any overfitting by feature selection, such that the proposed methods can be fairly compared with existing work in the literature. In addition, in order to avoid bias in the classification procedure, our proposed methods were evaluated based on the pure leave-one-women-out methodology.

8.3 Future Work

In this section, we outline the aspects which could be further investigated in future research on the basis of the work presented in this thesis.

- For breast density segmentation (see Chapter 3), one possible extension is to perform quantitative evaluation of the resulting segmentation according to radiologist-provided ground truth, obtained by manual annotation or using a user-interactive thresholding tool (Cumulus). For mammographic risk assessment, another possible extension is to further analyse the segmentation results and extract other geometric/topological features (e.g. compactness and Euler number) in addition to the area of the segmented regions, to link

the morphology of dense tissue to the BIRADS breast density classification.

- For local feature based mammographic tissue appearance modelling (see Chapter 4), the dimensionality of feature space built by some types of local features (e.g. LGA) is relatively higher. As indicated earlier, the resulting occurrence histograms of some local features have many empty or nearly empty histogram bins. The aspect of dimensionality reduction could be investigated in terms of noise removal and feature selection. Following on from this, one possible extension is to combine the dimensionality reduced histograms of various local features into an integrated histogram model, which means all local feature histograms derived from each mammogram constitute a new feature space. In addition, instead of extracting one local feature histogram from the whole breast region, a combination with the developed breast density segmentation could be investigated, by extracting multiple local feature histograms from the segmented sub-regions of different densities within the breast and using them to create the feature space.

- For blob based representation of mammographic parenchymal patterns (see Chapter 5), the Laplacian of Gaussian operator was used for blob detection, while alternative blob detectors, as described in Section 5.2, could be used. In the process of blob detection, one possible improvement is to investigate the aspect of false negative reduction in addition to removing false positives, in particular for the dense breasts indicating structureless and low contrast tissue. In addition, instead of only using the blob distribution over scales for mammographic risk assessment, greylevel/texture information within the resulting blobs could be incorporated to construct a composite representation of parenchymal patterns. Another possible area is to extend the 2D blob models of breast tissue into the 3D breast images by means of detecting 3D blobs (spheres or ellipsoids) from breast image volumes.

- We used a leave-one-woman-out cross-validation methodology in this thesis, where the distribution of training samples in the four BIRADS classes was unbalanced. BIRADS IV was not well represented in the training set due to its relatively smaller proportion in the database, which might partly explain the worse performance for this class. The unbalanced training sample problem could be further investigated to improve the classification.

- For topology based classification of microcalcification clusters (see Chapter 7), in addition to the connectivity, other properties of the resulting microcalcification graphs could be investigated, in order to generate a more sophisticated representation of microcalcification clusters. Features derived

from individual microcalcifications (e.g. shape factors) and/or the whole cluster (e.g. statistical cluster features) could also be incorporated to build a complete framework for microcalcification classification. In addition, instead of using feature vectors generated from microcalcification graphs to classify malignant and benign microcalcifications, a similarity measure between the microcalcification graphs could be defined, such that the classification can be implemented using the graphs directly without extracting topological features. Another extension will be to evaluate the developed method using a larger collection of digital mammograms.

- We used a kNN based classifier in this thesis for both breast density and microcalcification classification, which is a combination of the classic kNN classification and a distance weighted approach (as described in Section 6.2.1). A possible extension could be employing advanced classifiers or classifier ensembles (e.g. ANN, SVM, and random forests) for the classification.

- Finally, the mammographic image analysis methods presented in this thesis are developed mainly using digitised mammograms. As such another extension would be applying the proposed methods to full field digital mammograms and investigating the possible improvement.

8.4 Summary

In this chapter, we analysed and compared all the mammographic image analysis methods developed in this thesis. A comparison among the proposed methods was first presented. The classification results obtained by the individual methods were compared, and the reasons for the different performance of these methods were discussed. Subsequently, a complete comparison of the developed methods with related publications in the literature was provided. The aspects of experimental data, evaluation methodology for training and testing, classification algorithm, density classification scheme, and classification accuracy in the various approaches were discussed in detail. It was pointed out that the risk of overfitting or any possible bias should be avoided in the classification procedure to guarantee the feasibility of the developed methods in practice. Therefore, a high classification accuracy does not necessarily indicate a superior method if there is bias in the method. Finally, possible areas for future work were discussed. The conclusions drawn from the work presented in this thesis and the main contributions will be summarised in the following chapter.

Chapter 9

Conclusions

In this final chapter, we present the summary of this thesis and the conclusions drawn from the presented work. Moreover, we summarise the main contributions and novel aspects of our research in the field of mammographic image analysis. At the end of this chapter, a list of our publications related to this thesis is provided.

9.1 Summary of the Thesis

In this thesis, we have developed a set of mammographic image analysis methods using image processing and computer vision techniques, which constitute a complete framework for automated mammographic risk assessment, including breast region segmentation, breast density segmentation, parenchymal patterns modelling, and breast density classification. In addition, a novel method for automated microcalcification classification has been proposed, which aims to automatically discriminate between malignant and benign microcalcification clusters.

We started with a brief introduction of the research topics under investigation in this thesis, covering breast cancer, mammography, mammographic risk assessment, microcalcifications in mammograms, computer-aided diagnosis in mammography, and clinical utility of the present research. Following on from this, we carried out a literature review of mammographic image analysis and studied existing approaches in the field, which established a basis for our work performed in this thesis. From this review, we concluded that it is still well worth developing effective and robust mammographic image analysis methods for an actual implementation of automated mammographic risk assessment, although numerous methods have been developed in the past decade. In order to build a complete framework for mammographic risk assessment, we first developed a method for segmenting the breast region from mammograms as a pre-processing step. Subsequently, we developed a variety of methods for analysing mammographic images in terms of

estimating mammographic density and modelling mammographic parenchymal patterns, which were finally applied to mammographic risk assessment based on the BIRADS density classification.

For breast region segmentation (Chapter 3), the developed method was based on a combination of some well-known segmentation approaches, including histogram thresholding, multiscale edge detection, active contour model, and region growing. The performance of the developed method was tested using the EPIC and MIAS databases. A high percentage of acceptable segmentation results were obtained for both the breast-background segmentation (98.4% for EPIC and 100% for MIAS) and the pectoral muscle segmentation (93.5% for EPIC and 97.8% for MIAS). A qualitative comparison of the obtained segmentation results with those reported in related publications was provided, and moreover a direct comparison with two methods was made based on the MIAS database. It was demonstrated that the developed breast region segmentation method can provide sufficient performance for subsequent analysis in computer-aided diagnosis schemes.

For mammographic density estimation (Chapter 3), we developed two breast density segmentation methods based on a modified fuzzy-c means (MFCM) algorithm and a topographic approach, respectively. In the MFCM algorithm, spatial information was incorporated into the classic FCM clustering algorithm using an adaptive local window filter. Moreover, instead of performing clustering using every single pixel in mammographic images, a fast clustering procedure was implemented based on intensity histograms. To demonstrate the robustness of the MFCM algorithm to noise, it was extensively tested using a set of synthetic images corrupted by a variety of noise. The experimental results indicated that it outperformed recently published algorithms and showed high robustness/effectiveness with respect to different types of noise. When applying the MFCM algorithm to breast density segmentation, reasonable segmentation results were obtained corresponding to different density categories. In the topographic approach, a topographic map was exploited to represent a series of regions having a range of densities. A novel way was proposed to define dense tissue regions in the topographic map, in terms of *prominency* and *independency*. The resulting density maps indicated the area of dense tissue and the average intensity of each dense tissue region. High classification accuracies were obtained for the MIAS and DDSM databases (76.01% for MIAS and 81.23% for DDSM) according to the four BIRADS categories. In addition, the proposed breast density measure ($density = dense\ area \times average\ density$) was demonstrated to have the potential for quantifying breast density and tracking the quantitative change in breast density over time. A strong positive correlation was indicated for the 160 pairs of left/right MLO mammograms in the MIAS database (Pearson $r = 0.85$, Spearman $rho = 0.85$) as well as the 831 pairs of right MLO/CC

mammograms in the DDSM database (Pearson $r = 0.88$, Spearman $rho = 0.89$). For mammographic parenchymal pattern modelling (Chapter 4 and Chapter 5), we first proposed a method to describe breast tissue appearance in mammograms using local feature based texture representations, which were originally used for texture classification. Mammographic tissue was modelled based on statistical analysis of local appearance. Five strategies were investigated by employing different types of local features, including the aspects of intensity, texture, and geometry in the local neighbourhoods. A visual dictionary of local features was generated, which contained generic descriptions of local tissue appearance in mammograms. The overall mammographic tissue patterns were represented as occurrence histograms of local features in the visual dictionary. The validity of the proposed method was qualitatively and quantitatively evaluated using the MIAS database. For qualitative evaluation, every breast tissue pixel in each mammogram was labelled with the corresponding local feature, and the resulting label maps indicated various representations with respect to specific local tissue appearance. For quantitative evaluation, the resulting occurrence histograms were applied to breast density classification according to the four BIRADS categories. The highest performance (74.44%) was obtained by the approach based on the image-patch textons (Texton II) among the five strategies. An overall classification accuracy of 77.88% was obtained when combining the five individual classifications using a weighted average rule. In addition, we proposed a new multiscale blob based representation of mammographic parenchymal patterns. We focused on approximately blob-like tissue patterns in mammograms and used a set of multiscale blobs to represent the overall tissue patterns. The qualitative relations among blobs were incorporated to merge the overlapping blobs. A breast density measure ($BLOB_{norm}$) was defined and computed using the retained blobs. The validity of the proposed blob based representation was evaluated using the MIAS and DDSM databases. High classification accuracies (79.44% for MIAS and 76.90% for DDSM) were achieved for the four-class BIRADS density classification. A strong positive correlation (Pearson $r = 0.85$, Spearman $rho = 0.87$) was indicated for the defined breast density measure between the 160 pairs of left/right MLO mammograms in the MIAS database.

An overall evaluation of all the proposed methods described earlier was presented in Chapter 6, investigating the potential of these methods for mammographic risk assessment. Moreover, a comprehensive analysis on the obtained results was provided in Chapter 8, including an internal comparison among the proposed methods and a complete comparison with related publications in the literature. In addition, possible areas for future research were discussed at the end of Chapter 8. For microcalcification classification (Chapter 7), we developed a method for clas-

sifying malignant and benign microcalcification clusters based on morphological topology analysis. The connectivity/topology of microcalcifications was analysed to discriminate malignant from benign cases. Instead of extracting features at a single scale, a multiscale topological feature vector was extracted from a set of constructed microcalcification graphs, describing the topological structure of microcalcification clusters at multiple scales. The validity of the proposed method was evaluated using three datasets: MIAS, DDSM, and a full-field digital dataset. Four k-Nearest Neighbours based algorithms (kNN, FNN, FRNN and VQNN) were employed for the classification. High classification accuracies (up to 96%) and good ROC results (area under the ROC curve up to 0.97) were achieved. A full comparison with relate publications was also provided.

In summary, we conclude that the results obtained so far are very promising, which indicates that the methods presented in this thesis have the potential for automated and accurate mammographic risk assessment as well as microcalcification classification. When comparing the proposed methods, the methods focusing on dense tissue detection/segmentation in the breast (i.e. the topographic representation based method for breast density segmentation and the multiscale blob based method for parenchymal patterns representation) seem better suited to the BIRADS density classification and tend to produce better classification results, compared to the methods modelling breast tissue within the whole breast (i.e. MFCM, LBP, LGA, BIR, Texton I and Texton II). Intensity information plays an important role in breast tissue density characterisation/classification, and improved performance can be achieved when combined with spatial (texture) information (MFCM achieved better performance when combined with Texton II). An integration of the various methods, such as combining multiple classifiers (i.e. combining the five classifiers built by LBP, LGA, BIR, Texton I and Texton II), concatenating features extracted using different approaches, or concatenating features derived from MLO/CC view mammograms, tends to provide better performance than using the methods individually. Moreover, for the larger dataset taken from the DDSM database, we performed the stratified 10-fold cross-validation in addition to leave-one-woman-out and equivalent classification results were obtained between the two evaluation methodologies.

9.2 Contributions and Novel Aspects

The main contributions and novel aspects of this thesis are summarised as follows:

- An extensive literature review of mammographic image analysis was presented, which includes mammographic image segmentation, mammographic

density estimation, mammographic parenchymal patterns characterisation, and breast density classification. The main points of all the reviewed approaches were described and a summary table was provided for each area, which enables a quick reference to the key techniques in these approaches.

- A new method for breast region segmentation was developed, which has been effectively applied to the EPIC, MIAS, and DDSM databases. The obtained segmentation results provide a fundamental basis for subsequent analysis.

- A modified fuzzy c-means algorithm was developed, which incorporates spatial information into the classic fuzzy c-means algorithm to reduce the sensitivity to noise and intensity inhomogeneity. Moreover, a fast clustering procedure was performed by grouping the pixels with the same intensity value in each iteration step to improve computational efficiency. This enables this algorithm to be applied to large size mammographic images for segmenting the breast region into various density sub-regions.

- A new method for detecting/segmenting dense tissue regions was developed based on the topographic approach. A *shape* tree was constructed to describe the topological and geometrical structure of the connected components within the topographic map. A novel way was proposed to define dense tissue regions as *prominent/independent shapes* in the topographic map. This is a first attempt to model mammographic density using topography and detect/segment dense tissue regions based on a tree representation.

- A thorough investigation of using various local features for modelling mammographic tissue appearance was conducted. A qualitative and quantitative comparison was made between different local features using the resulting label maps and occurrence histograms. The resulting label maps provided a useful way of understanding breast tissue appearance with respect to intensity, texture, and geometry.

- A new representation of mammographic parenchymal patterns was proposed. Instead of modelling breast tissue patterns formed by all tissue types, we focused on dense tissue with approximately blob-like structures. Blob-like dense tissue regions were extracted by detecting bright blobs from the breast region, and as such mammographic parenchymal patterns were represented with a set of multiscale blobs. This is a first attempt to model parenchymal patterns by focusing on dense tissue with blob-like structures.

- A new method for the classification of microcalcification clusters was developed based on topological analysis. A microcalcification graph was con-

structed to represent the topological structure of microcalcification clusters. The connectivity between individual microcalcifications was analysed to classify microcalcification clusters into malignant and benign. This is distinct from existing approaches which tend to focus on the shape of individual microcalcifications and/or global (statistical) cluster features.

- A complete evaluation of the developed methods was conducted using well-known mammogram databases in the field of mammographic image analysis. Moreover, a complete comparison was made between the developed methods as well as with related publications in the literature.

- Finally, a framework for automated mammographic risk assessment has been developed, which has the potential for computer-aided diagnosis.

9.3 Publication List

A list of publications resulting from this thesis is shown as follows:

1. [**MIUA2010**] Z. Chen and R. Zwiggelaar. Segmentation of the Breast Region with Pectoral Muscle Removal in Mammograms. *In Proceedings of the* 14^{th} *Medical Image Understanding and Analysis*, pp. 71-75, Coventry, UK, 06-07 July, 2010.

2. [**ITAB2010**] Z. Chen and R. Zwiggelaar. A Modified Fuzzy C-Means Algorithm for Breast Tissue Density Segmentation in Mammograms. *In Proceedings of the* 10^{th} *International Conference on Information Technology and Applications in Biomedicine*, Corfu, Greece, 03-05 November, 2010.

3. [**MIUA2011**] Z. Chen, E. Denton, and R. Zwiggelaar. Modelling Breast Tissue in Mammograms for Mammographic Risk Assessment. *In Proceedings of the* 15^{th} *Medical Image Understanding and Analysis*, pp. 37-41, London, UK, 14-15 July, 2011.

4. [**MICCAI2011**] Z. Chen, E. Denton, and R. Zwiggelaar. A Multi-Scale Blob Based Representation of Mammographic Parenchymal Patterns for Breast Density Classification. *In Proceedings of MICCAI Workshop on Breast Image Analysis*, pp. 169-176, Toronto, Canada, 18-22, September, 2011.

5. [**BMEI2011**] Z. Chen, E. Denton, and R. Zwiggelaar. Local Feature Based Mammographic Tissue Pattern Modelling and Breast Density Classification. *In Proceedings of the* 4^{th} *International Conference on BioMedical Engineering and Informatics*, pp. 351-355, Shanghai, China, 15-17 October, 2011.

6. [**IWDM2012**] Z. Chen, E. Denton, and R. Zwiggelaar. Classification of Microcalcification Clusters Based on Morphological Topology Analysis. Lecture Notes in Computer Science, vol. 7361, pp. 521-528, 2012.

7. [**ICIP2012**] Z. Chen, E. Denton, and R. Zwiggelaar. Topographic Representation Based Breast Density Segmentation for Mammographic Risk Assessment. *In Proceedings of the* 19^{th} *International Conference on Image Processing*, pp. 1993-1996, Orlando, Florida, USA, 30 September-03 October, 2012.

8. [**MIUA2012**] Z. Chen, A. Oliver, E. Denton, and R. Zwiggelaar. Classification of Microcalcification Clusters Using Topological Structure Features. *In Proceedings of the* 16^{th} *Medical Image Understanding and Analysis*, pp. 37-42, Swansea, UK, 09-11 July, 2012.

9. [**MIUA2012**] Z. Chen, E. Denton, and R. Zwiggelaar. A Statistical Approach to Modelling Breast Tissue Appearance in Mammograms. *In Proceedings of the* 16^{th} *Medical Image Understanding and Analysis*, pp. 43-48, Swansea, UK, 09-11 July, 2012.

10. [**BMEI2012**] Z. Chen and R. Zwiggelaar. A Combined Method for Automatic Identification of the Breast Boundary in Mammograms. *In Proceedings of the* 5^{th} *International Conference on BioMedical Engineering and Informatics*, pp. 123-127, Chongqing, China, 16-18 October, 2012.

11. [**IbPRIA2013**] Z. Chen, A. Oliver, E. Denton, and R. Zwiggelaar. Automated Mammographic Risk Classification Based on Breast Density Estimation. Lecture Notes in Computer Science, vol. 7887, pp. 237-244, 2013.

12. [**BMVA2013**] Z. Chen, E. Denton, and R. Zwiggelaar. Local Feature Based Breast Tissue Appearance Modelling for Mammographic Risk Assessment. *Annals of the British Machine Vision Association*, 2013(3), 1-19.

Bibliography

Abdel-Mottaleb, M., Carman, C.S., Hill, C.R., & Vafai, S. 1996. Locating the boundary between the breast skin edge and the background in digitized mammograms. *Pages 467–470 of: the 3rd International Workshop on Digital Mammography.*

Adel, M., Rasigni, M., Bourennane, S., & Juhan, V. 2007. Statistical segmentation of regions of interest on a mammographic image. *EURASIP Journal on Advances in Signal Processing*, **2007**(2).

Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag, A.A., & Moriarty, T. 2002. A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging*, **21**(3), 193–199.

Ahuja, N. 1982. Dot pattern processing using voronoi neighborhoods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 336–343.

American Cancer Society. 2011a. *Breast Cancer Facts & Figures 2011–2012.* Atlanta: American Cancer Society.

American Cancer Society. 2011b. *Global Cancer Facts & Figures.* 2 edn. Atlanta: American Cancer Society.

American College of Radiology. 1998. *Illustrated Breast Imaging Reporting and Data System BIRADS.* 3 edn. Philadelphia, PA: American College of Radiology.

Amir, E., Freedman, O.C., Seruga, B., & Evans, D.G. 2010. Assessing women at high risk of breast cancer: a review of risk assessment models. *Journal of the National Cancer Institute*, **102**(10), 680–691.

Arivazhagan, S., Ganesan, L., & Priyal, S.P. 2006. Texture classification using Gabor wavelets based rotation invariant features. *Pattern Recognition Letters*, **27**(16), 1976–1982.

Babaud, J., Witkin, A.P., Baudin, M., & Duda, R.O. 1986. Uniqueness of the gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(1), 26–33.

Bakic, P., Brzakovic, D., Brzakovic, P., & Zhu, Z. 1998. An approach to using a generalized breast model to segment digital mammograms. *Pages 84–89 of: the 11th IEEE Symposium on Computer-Based Medical Systems.*

Bakic, P.R., Albert, M., Brzakovic, D., & Maidment, A.D.A. 2002a. Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation. *Medical Physics*, **29**(9), 2131–2139.

Bakic, P.R., Albert, M., Brzakovic, D., & Maidment, A.D.A. 2002b. Mammogram synthesis using a 3D simulation. II. Evaluation of synthetic mammogram texture. *Medical Physics*, **29**(9), 2140–2151.

Bakic, P.R., Carton, A.K., Kontos, D., Zhang, C., Troxel, A.B., & Maidment, A.D.A. 2009. Breast Percent Density: Estimation on Digital Mammograms and Central Tomosynthesis Projections. *Radiology*, **252**(1), 40–49.

Bay, H., Tuytelaars, T., & Van Gool, L. 2006. SURF: Speeded Up Robust Features. *Pages 404–417 of: Leture Notes in Computer Science*, vol. 3951.

Beck, J.R., & Schultz, E.K. 1986. The use of relative operating characteristic (ROC) curves in test performance evaluation. *Archives of Pathology & Laboratory Medicine*, **110**, 13–20.

Berg, W.A., Campassi, C., Langenberg, P., & Sexton, M.J. 2000. Breast Imaging Reporting and Data System: Inter-and Intraobserver Variability in Feature Analysis and Final Assessment. *American Journal of Roentgenology*, **174**(6), 1769–1777.

Berry, D.A., Cronin, K.A., Plevritis, S.K., Fryback, D.G., Clarke, L., Zelen, M., Mandelblatt, J.S., Yakovlev, A.Y., Habbema, J.D.F., & Feuer, E.J. 2005. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine*, **353**(17), 1784–1792.

Betal, D., Roberts, N., & Whitehouse, G.H. 1997. Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology. *British Journal of Radiology*, **70**(837), 903–917.

Bezdek, J.C. 1981. *Pattern recognition with fuzzy objective function algorithms.* New York: Plenum Press.

Bick, U., Giger, M.L., Schmidt, R.A., Nishikawa, R.M., Wolverton, D.E., & Doi, K. 1995. Automated segmentation of digitized mammograms. *Academic Radiology*, **2**(1), 1–9.

Bilgin, C., Demir, C., Nagi, C., & Yener, B. 2007. Cell-graph mining for breast tissue modeling and classification. *Pages 5311–5314 of: the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*

Bilgin, C., Bullough, P., Plopper, G.E., & Yener, B. 2010. ECM-aware cell-graph mining for bone tissue modeling and classification. *Data Mining and Knowledge Discovery*, **20**(3), 416–438.

Bishop, C.M. 1995. *Neural networks for pattern recognition.* Oxford: Oxford University Press.

Bishop, C.M. 2006. *Pattern recognition and machine learning.* New York: Springer.

Blot, L., & Zwiggelaar, R. 2001. Background texture extraction for the classification of mammographic parenchymal patterns. *Pages 145–148 of: the 5th Medical Image Understanding and Analysis.*

Blot, L., & Zwiggelaar, R. 2005. A volumetric approach to glandularity estimation in mammography: a feasibility study. *Physics in Medicine and Biology*, **50**, 695–708.

Boehm, H.F., Schneider, T., Buhmann-Kirchhoff, S.M., Schlossbauer, T., Rjosk-Dendorfer, D., Britsch, S., & Reiser, M. 2008. Automated Classification of Breast Parenchymal Density: Topologic Analysis of X-Ray Attenuation Patterns Depicted with Digital Mammography. *American Journal of Roentgenology*, **191**(6), W275–W282.

Bovik, A.C., Clark, M., & Geisler, W.S. 1990. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(1), 55–73.

Bovis, K., & Singh, S. 2002. Classification of mammographic breast density using a combined classifier paradigm. *Pages 177–180 of: the 6th Medical Image Understanding and Analysis.*

Boyd, N., Martin, L., Gunasekara, A., Melnichouk, O., Maudsley, G., Peressotti, C., Yaffe, M., & Minkin, S. 2009. Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes. *Cancer Epidemiology Biomarkers & Prevention*, **18**(6), 1754–1762.

Boyd, N.F., Byng, J.W., Jong, R.A., Fishell, E.K., Little, L.E., Miller, A.B., Lockwood, G.A., Tritchler, D.L., & Yaffe, M.J. 1995. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian

National Breast Screening Study. *Journal of the National Cancer Institute*, **87**(9), 670–675.

Boyd, N.F., Lockwood, G.A., Byng, J.W., Tritchler, D.L., & Yaffe, M.J. 1998. Mammographic densities and breast cancer risk. *Cancer Epidemiology Biomarkers & Prevention*, **7**(12), 1133.

Boyd, N.F., Guo, H., Martin, L.J., Sun, L., Stone, J., Fishell, E., Jong, R.A., Hislop, G., Chiarelli, A., Minkin, S., & Yaffe, M.J. 2007. Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine*, **356**(3), 227–236.

Brink, A.D., & Pendock, N.E. 1996. Minimum cross-entropy threshold selection. *Pattern Recognition*, **29**(1), 179–188.

Bruce, V., Green, P.R., & Georgeson, M.A. 1996. *Visual Perception: Physiology, Psychology, and Ecology*. Exeter, UK: Psychology Press.

Bueno, G., Vállez, N., Déniz, O., Esteve, P., Rienda, M.A., Arias, M., & Pastor, C. 2011. Automatic breast parenchymal density classification integrated into a CADe system. *International Journal of Computer Assisted Radiology and Surgery*, **6**(3), 309–318.

Burt, P., & Adelson, E. 1983. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, **31**(4), 532–540.

Byng, J.W., Boyd, N.F., Fishell, E., Jong, R.A, & Yaffe, M.J. 1994. The quantitative analysis of mammographic densities. *Physics in Medicine and Biology*, **39**, 1629–1638.

Byng, J.W., Boyd, N.F., Fishell, E., Jong, R.A., & Yaffe, M.J. 1996a. Automated analysis of mammographic densities. *Physics in Medicine and Biology*, **41**, 909–923.

Byng, J.W., Boyd, N.F., Little, L., Lockwood, G., Fishell, E., Jong, R.A., & Yaffe, M.J. 1996b. Symmetry of projection in the quantitative analysis of mammographic images. *European Journal of Cancer Prevention*, **5**(5), 319–327.

Byng, J.W., Yaffe, M.J., Jong, R.A., Shumak, R.S., Lockwood, G.A., Tritchler, D.L., & Boyd, N.F. 1998. Analysis of mammographic density and breast cancer risk from digitized mammograms. *Radiographics*, **18**(6), 1587–1598.

Cai, W., Chen, S., & Zhang, D. 2007. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern Recognition*, **40**(3), 825–838.

Caldwell, C.B., Stapleton, S.J., Holdsworth, D.W., Jong, R.A., Weiser, W.J., Cooke, G., & Yaffe, M.J. 1990. Characterisation of mammographic parenchymal pattern by fractal dimension. *Physics in Medicine and Biology*, **35**(2), 235–247.

Camilus, K.S., Govindan, V.K., & Sathidevi, P.S. 2010. Computer-aided identification of the pectoral muscle in digitized mammograms. *Journal of Digital Imaging*, **23**(5), 562–580.

Cancer Research UK. 2012a (November). http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/incidence/.

Cancer Research UK. 2012b (August). http://www.cancerresearchuk.org/cancer-help/type/breast-cancer/about/screening/who-is-screened-for-breast-cancer.

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(6), 679–698.

Caselles, V., Catté, F., Coll, T., & Dibos, F. 1993. A geometric model for active contours in image processing. *Numerische Mathematik*, **66**(1), 1–31.

Castella, C., Kinkel, K., Eckstein, M.P., Sottas, P.E., Verdun, F.R., & Bochud, F.O. 2007. Semiautomatic mammographic parenchymal patterns classification using multiple statistical features. *Academic Radiology*, **14**(12), 1486–1499.

Castellano, G., Bonilha, L., Li, L.M., & Cendes, F. 2004. Texture analysis of medical images. *Clinical Radiology*, **59**(12), 1061–1069.

Chan, H. P., Sahiner, B., Lam, K. L., Petrick, N., Helvie, M. A., Goodsitt, M. M., & Adler, D. D. 1998. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. *Medical Physics*, **25**(10), 2007–2019.

Chan, H.P., Sahiner, B., Petrick, N., Helvie, M.A., Lam, K.L., Adler, D.D., & Goodsitt, M.M. 1997. Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network. *Physics in Medicine and Biology*, **42**, 549–567.

Chandrasekhar, R., & Attikiouzel, Y. 1996. Gross segmentation of mammograms using a polynomial model. *Pages 1056–1058 of: the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine.*, vol. 3.

Chandrasekhar, R., & Attikiouzel, Y. 2000. Automatic breast border segmentation by background modeling and subtraction. *Pages 560–565 of: the 5th International Workshop on Digital Mammography.*

Chang, Y.H., Wang, X.H., Hardesty, L.A., Chang, T.S., Poller, W.R., Good, W.F., & Gur, D. 2002. Computerized assessment of tissue composition on digitized mammograms. *Academic Radiology*, **9**(8), 899–905.

Chaudhuri, B.B., & Sarkar, N. 1995. Texture segmentation using fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(1), 72–77.

Cheng, H.D., Cai, X., Chen, X., Hu, L., & Lou, X. 2003. Computer-Aided Detection and Classification of Microcalcifications in Mammograms: A Survey. *Pattern Recognition*, **36**(12), 2967–2991.

Cheng, Y. 1995. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(8), 790–799.

Chiu, S.H., & Liaw, J.J. 2005. An effective voting method for circle detection. *Pattern Recognition Letters*, **26**(2), 121–133.

Chuang, K.S., Tzeng, H.L., Chen, S., Wu, J., & Chen, T.J. 2006. Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics*, **30**(1), 9–15.

Ciatto, S., Houssami, N., Apruzzese, A., Bassetti, E., Brancato, B., Carozzi, F., Catarzi, S., Lamberini, M.P., Marcelli, G., Pellizzoni, R., Pesce, B., Risso, G., Russo, F., & Scorsolini, A. 2005. Categorizing breast mammographic density: intra-and interobserver reproducibility of BI-RADS density categories. *The Breast*, **14**(4), 269–275.

Cleveland, W.S., & Devlin, S.J. 1988. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 596–610.

Cohen, L.D. 1991. On active contour models and balloons. *CVGIP: Image Understanding*, **53**(2), 211–218.

Comaniciu, D., & Meer, P. 1999. Mean shift analysis and applications. *Pages 1197–1203 of: the 7th IEEE International Conference on Computer Vision*, vol. 2.

Conners, R.W., & Harlow, C.A. 1980. A theoretical comparison of texture algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 204–222.

Conte, D., Foggia, P., Sansone, C., & Vento, M. 2004. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, **18**(3), 265–298.

Cornelis, C., De Cock, M., & Radzikowska, A. 2007. Vaguely quantified rough sets. *Lecture Notes in Artificial Intelligence*, **4482**, 87–94.

Crosier, M., & Griffin, L.D. 2010. Using basic image features for texture classification. *International Journal of Computer Vision*, **88**(3), 447–460.

Cross, G.R., & Jain, A.K. 1983. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **5**(1), 25–39.

D'Astous, F., & Jernigan, M.E. 1984. Texture discrimination based on detailed measures of the power spectrum. *Pages 83–86 of: International Conference on Pattern Recognition*.

Daugman, J.G., *et al.* 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of Optical Society of America, A: Optics and Image Science*, **2**, 1160–1169.

Davies, E.R. 1997. *Machine vision: theory, algorithms, practicalities*. London: Academic Press.

Day, N., Oakes, S., Luben, R., Khaw, K.T., Bingham, S., Welch, A., & Wareham, N. 1999. EPIC-Norfolk: study design and characteristics of the cohort. *British Journal of Cancer*, **80**(Suppl 1), 95–103.

Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 1–38.

Dengler, J., Behrens, S., & Desaga, J.F. 1993. Segmentation of microcalcifications in mammograms. *IEEE Transactions on Medical Imaging*, **12**(4), 634–642.

Dhawan, A.P., Chitre, Y., & Kaiser-Bonasso, C. 1996. Analysis of Mammographic Microcalcifications Using Gray-level Image Structure Features. *IEEE Transactions on Medical Imaging*, **15**(3), 246–259.

Diestel, R. 2010. *Graph theory*. Fourth edn. Heidelberg: Springer-Verlag.

Doi, K. 2007. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, **31**(4-5), 198–211.

Duda, R. O., Hart, P. E., & Stork, D. G. 2001a. *Pattern classification*. Second edn. New York: John Wiley and Sons.

Duda, R.O., & Hart, P.E. 1973. *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.

Duda, R.O., Hart, P.E., & Stork, D.G. 2001b. *Pattern Classification*. New York: John Wiley & Sons.

Dunn, J.C. 1973. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, **3**, 32–57.

Egmont-Petersen, M., De Ridder, D., & Handels, H. 2002. Image processing with neural networks–a review. *Pattern Recognition*, **35**(10), 2279–2301.

Eurostat. 2009. *Health Statistics Atlas on Mortality in the European Union*. Office for Official Publications of the European Union.

Farag, A.A., El-Baz, A.S., & Gimel'farb, G. 2006. Precise segmentation of multimodal images. *IEEE Transactions on Image Processing*, **15**(4), 952–968.

Feig, S.A., Galkin, B.M., & Muir, H.D. 1987. Evaluation of breast microcalcifications by means of optically magnified tissue specimen radiographs. *Recent Results in Cancer Research*, **105**, 111–123.

Felzenszwalb, P.F., & Huttenlocher, D.P. 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision*, **59**(2), 167–181.

Ferrari, R.J., Rangayyan, R.M., Desautels, J.E.L., & Frere, A.F. 2000. Segmentation of mammograms: identification of the skin-air boundary, pectoral muscle, and fibro-glandular disc. *Pages 573–579 of: the 5th International Workshop on Digital Mammography*.

Ferrari, R.J., Rangayyan, R.M., Desautels, J.E.L., Borges, R.A., & Frere, A.F. 2004a. Automatic identification of the pectoral muscle in mammograms. *IEEE Transactions on Medical Imaging*, **23**(2), 232–245.

Ferrari, R.J., Frère, A.F., Rangayyan, R.M., Desautels, J.E.L., & Borges, R.A. 2004b. Identification of the breast boundary in mammograms using active contour models. *Medical and Biological Engineering and Computing*, **42**(2), 201–208.

Ferrari, R.J., Rangayyan, R.M., Borges, R.A., & Frere, A.F. 2004c. Segmentation of the fibro-glandular disc in mammogrms using Gaussian mixture modelling. *Medical and Biological Engineering and Computing*, **42**(3), 378–387.

Freer, T.W., & Ulissey, M.J. 2001. Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center. *Radiology*, **220**(3), 781–786.

Galloway, M.M. 1975. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, **4**(2), 172–179.

Geisser, S. 1993. *Predictive inference: An introduction*. New York: Chapman & Hall.

Gong, Y., Brady, M., & Petroudi, S. 2006. Texture based mammogram classification and segmentation. *Pages 616–625 of: the 8th International Workshop on Digital Mammography*.

Gonzalez, R.C., & Woods, R.E. 2002. *Digital Image Processing*. 2 edn. Prentice Hall Press.

Gonzalez, R.C., Woods, R.E., & Eddins, S.L. 2004. *Digital image processing using MATLAB*. Pearson Prentice Hall.

Gøtzsche, P.C., & Nielsen, M. 2009. Screening for breast cancer with mammography. *The Cochrane Library*, 1–48.

Greif, J.M. 2010. Mammographic screening for breast cancer: An invited review of the benefits and costs. *The Breast*, **19**(4), 268–272.

Gunduz, C., Yener, B., & Gultekin, S. H. 2004. The cell graphs of cancer. *Bioinformatics*, **20**(Suppl 1), i145–i151.

Guo, C., Zhu, S.C., & Wu, Y.N. 2003. Towards a mathematical theory of primal sketch and sketchability. *Pages 1228–1235 of: 9th IEEE International Conference on Computer Vision*.

Gur, D., Sumkin, J.H., Rockette, H.E., Ganott, M., Hakim, C., Hardesty, L., Poller, W.R., Shah, R., & Wallace, L. 2004. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *Journal of the National Cancer Institute*, **96**(3), 185–190.

Hadley, E., Denton, E., & Zwiggelaar, R. 2006. Mammographic risk assessment based on anatomical linear structures. *Pages 626–633 of: the 8th International Workshop on Digital Mammography*.

Haji, M. M., & Katebi, S. D. 2005. An efficient text segmentation technique based on naive Bayes classifier. *ICGST International Journal on Graphics, Vision and Image Processing*, **5**, 21–30.

Halkiotis, S., Botsis, T., & Rangoussi, M. 2007. Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks. *Signal Processing*, **87**(7), 1559–1568.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. 2009. The WEKA data mining software: an update. *SIGKDD Explorations*, **11**(1), 10–18.

Haralick, R.M. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE*, **67**(5), 786–804.

Haralick, R.M., Shanmugam, K., & Dinstein, I.H. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, **3**(6), 610–621.

Haris, K., Efstratiadis, S.N., Maglaveras, N., & Katsaggelos, A.K. 1998. Hybrid image segmentation using watersheds and fast region merging. *IEEE Transactions on Image Processing*, **7**(12), 1684–1699.

Hartman, K., Highnam, R., Warren, R., & Jackson, V. 2008. Volumetric assessment of breast tissue composition from FFDM images. *Pages 33–39 of: the 9th International Workshop on Digital Mammography.*

Harvey, J.A., & Bovbjerg, V.E. 2004. Quantitative Assessment of Mammographic Breast Density: Relationship with Breast Cancer Risk. *Radiology*, **230**(1), 29–41.

Hassanien, A.E. 2007. Fuzzy rough sets hybrid scheme for breast cancer detection. *Image and Vision Computing*, **25**(2), 172–183.

He, W., Muhimmah, I., Denton, E., & Zwiggelaar, R. 2008. Mammographic segmentation based on texture modelling of tabár mammographic building blocks. *Pages 17–24 of: the 9th International Workshop on Digital Mammography.* Springer.

He, W., Denton, E., & Zwiggelaar, R. 2010. Mammographic Image Segmentation and Risk Classification Using a Novel Texture Signature Based Methodology. *Pages 526–533 of: the 10th International Workshop on Digital Mammography.*

He, W., Denton, E.R.E., Stafford, K., & Zwiggelaar, R. 2011. Mammographic image segmentation and risk classification based on mammographic parenchymal patterns and geometric moments. *Biomedical Signal Processing and Control*, 321–329.

He, W., Denton, E., & Zwiggelaar, R. 2012. Mammographic segmentation and risk classification using a novel binary model based bayes classifier. *Pages 40–47 of: the 11th International Workshop on Digital Mammography.*

Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, P. 2000. The digital database for screening mammography. *Pages 212–218 of: the 5th International Workshop on Digital Mammography.*

Heine, J.J., & Velthuizen, R.P. 2000. A statistical methodology for mammographic density detection. *Medical Physics*, **27**, 2644–2651.

Heine, J.J., Kallergi, M., Chetelat, S.M., & Clarke, L.P. 1998. Multiresolution wavelet approach for separating the breast region from the background in high resolution digital mammography. *Computional Imaging and Vision*, **13**, 295–298.

Heine, J.J., Carston, M.J., Scott, C.G., Brandt, K.R., Wu, F.F., Pankratz, V.S., Sellers, T.A., & Vachon, C.M. 2008. An automated approach for estimation of breast density. *Cancer Epidemiology Biomarkers & Prevention*, **17**(11), 3090–3097.

Highnam, R., & Brady, M. 1999. *Mammographic image analysis.* Kluwer Academic Publishers.

Highnam, R., Pan, X., Warren, R., Jeffreys, M., Davey Smith, G., & Brady, M. 2006. Breast composition measurements using retrospective standard mammogram form (SMF). *Physics in Medicine and Biology*, **51**, 2695–2713.

Hong, B.W., & Brady, M. 2003. A topographic representation for mammogram segmentation. *Pages 730–737 of: the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention.*

Hong, B.W., & Sohn, B.S. 2010. Segmentation of regions of interest in mammograms in a topographic approach. *IEEE Transactions on Information Technology in Biomedicine*, **14**(1), 129–139.

Hough, P.V.C. 1962. *Method and means for recognizing complex patterns.* US Patent 3,069,654.

Hsiao, J.Y., & Sawchuk, A.A. 1989. Supervised textured image segmentation using feature smoothing and probabilistic relaxation techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(12), 1279–1292.

Huo, Z., Giger, M.L., Wolverton, D.E., Zhong, W., Cumming, S., & Olopade, O.I. 2000. Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection. *Medical Physics*, **27**, 4–12.

Idrissa, M., & Acheroy, M. 2002. Texture classification using Gabor filters. *Pattern Recognition Letters*, **23**(9), 1095–1102.

Jackway, P.T. 2000. Improved morphological top-hat. *Electronics Letters*, **36**(14), 1194–1195.

Jain, A.K., & Farrokhnia, F. 1991. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, **24**(12), 1167–1186.

Jain, A.K., Murty, M.N., & Flynn, P.J. 1999. Data clustering: a review. *ACM Computing Surveys (CSUR)*, **31**(3), 264–323.

Jensen, R., & Cornelis, C. 2011. Fuzzy rough nearest neighbour classification and prediction. *Theoretical Computer Science*, **412**(42), 5871–5884.

Jensen, R., & Shen, Q. 2009. New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems*, **17**(4), 824–838.

Jiang, Y., Nishikawa, R.M., Wolverton, D.E., Metz, C.E., Schmidt, R.A., & Doi, K. 1997. Computerized classification of malignant and benign clustered microcalcifications in mammograms. *Pages 521–523 of: 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*

Jojic, N., Frey, B.J., & Kannan, A. 2003. Epitomic analysis of appearance and shape. *Pages 34–41 of: 9th IEEE International Conference on Computer Vision.*

Kadir, T., & Brady, M. 2001. Saliency, scale and image description. *International Journal of Computer Vision*, **45**(2), 83–105.

Kalager, M., Zelen, M., Langmark, F., & Adami, H.O. 2010. Effect of screening mammography on breast-cancer mortality in Norway. *New England Journal of Medicine*, **363**(13), 1203–1210.

Kallenberg, M.G.J., Lokate, M., van Gils, C.H., & Karssemeijer, N. 2011. Automatic breast density segmentation: an integration of different approaches. *Physics in Medicine and Biology*, **56**, 2715–2729.

Kang, J., Min, L., Luan, Q., Li, X., & Liu, J. 2009. Novel modified fuzzy c-means algorithm with applications. *Digital Signal Processing*, **19**(2), 309–319.

Karssemeijer, N. 1998. Automated classification of parenchymal patterns in mammograms. *Physics in Medicine and Biology*, **43**, 365–378.

Kass, M., Witkin, A., & Terzopoulos, D. 1988. Snakes: Active contour models. *International Journal of Computer Vision*, **1**(4), 321–331.

Keller, B., Nathan, D., Wang, Y., Zheng, Y., Gee, J., Conant, E., & Kontos, D. 2011. Adaptive multi-cluster fuzzy c-means segmentation of breast parenchymal tissue in digital mammography. *Pages 562–569 of: the 14th International Conference on Medical Image Computing and Computer Assisted Intervention*.

Keller, B. M., Nathan, D. L., Wang, Y., Zheng, Y., Gee, J. C., Conant, E. F., & Kontos, D. 2012. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. *Medical Physics*, **39**(8), 4903–4917.

Keller, J.M., Gray, M.R., & Givens, J.A. 1985. A fuzzy K-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, **15**(4), 580–585.

Keller, J.M., Chen, S., & Crownover, R.M. 1989. Texture description and segmentation through fractal geometry. *Computer Vision, Graphics, and Image Processing*, **45**(2), 150–166.

Kinoshita, S.K., De Azevedo-Marques, P.M., Pereira, R.R., Rodrigues, J.A.H., & Rangayyan, R.M. 2007. Content-based retrieval of mammograms using visual features related to breast density patterns. *Journal of Digital Imaging*, **20**(2), 172–190.

Kittler, J. 1986. *Feature selection and extraction*. Handbook of Pattern Recognition and Image Processing. New York: Academic.

Kittler, J., & Illingworth, J. 1986. Minimum error thresholding. *Pattern Recognition*, **19**(1), 41–47.

Koenderink, J.J. 1984. The structure of images. *Biological Cybernetics*, **50**(5), 363–370.

Kolb, T.M., Lichy, J., & Newhouse, J.H. 2002. Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations. *Radiology*, **225**(1), 165–175.

Kwok, S.M., Chandrasekhar, R., Attikiouzel, Y., & Rickard, M.T. 2004. Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. *IEEE Transactions on Medical Imaging*, **23**(9), 1129–1140.

Lao, Z., & Huo, Z. 2009. Quantitative assessment of breast dense tissue on mammograms. *Pages 2605–2608 of: the 16th IEEE International Conference on Image Processing.*

Lau, T.K., & Bischof, W.F. 1991. Automated detection of breast tumors using the asymmetry approach. *Computers and Biomedical Research*, **24**(3), 273–295.

Laws, K.I. 1980. *Textured image segmentation.* Ph.D. thesis, Image Processing Institute, University of Southern California.

Li, H., Giger, M.L., Olopade, O.I., Margolis, A., Lan, L., & Chinander, M.R. 2005. Computerized Texture Analysis of Mammographic Parenchymal Patterns of Digitized Mammograms. *Academic Radiology*, **12**(7), 863–873.

Li, H., Giger, M.L., Olopade, O.I., & Chinander, M.R. 2008. Power spectral analysis of mammographic parenchymal patterns for breast cancer risk assessment. *Journal of Digital Imaging*, **21**(2), 145–152.

Li, H., Giger, M.L., Lan, L., Brown, J.B., MacMahon, A., Mussman, M., Olopade, O.I., & Sennett, C. 2012. Computerized Analysis of Mammographic Parenchymal Patterns on a Large Clinical Dataset of Full-Field Digital Mammograms: Robustness Study with Two High-Risk Datasets. *Journal of Digital Imaging*, 591–598.

Li, S. Z. 2009. *Markov random field modeling in image analysis.* third edn. New York: Springer-Verlag.

Lindeberg, T. 1994. *Scale-Space Theory in Computer Vision.* Netherlands: Kluwer Academic Publishers/Springer.

Lindeberg, T. 1998. Feature detection with automatic scale selection. *International Journal of Computer Vision*, **30**(2), 79–116.

Lindeberg, T. 2008. *Pages 2495–2504 of: Encyclopedia of Computer Science and Engineering*, vol. IV.

Liu, J., White, J.M., & Summers, R.M. 2010. Automated detection of blob structures by hessian analysis and object scale. *Pages 841–844 of: 17th IEEE International Conference on Image Processing.*

Lloyd, S. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, **28**(2), 129–137.

Lou, S.L., Lin, H.D., Lin, K.P., & Hoogstrate, D. 2000. Automatic breast region extraction from digital mammograms for PACS and telemammography applications. *Computerized Medical Imaging and Graphics*, **24**(4), 205–220.

Lowe, D.G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, **60**(2), 91–110.

Lu, L.J.W., Nishino, T.K., Khamapirad, T., Grady, J.J., Leonard, M.H., & Brunder, D.G. 2007. Computing mammographic density from a multiple regression model constructed with image-acquisition parameters from a full-field digital mammographic unit. *Physics in Medicine and Biology*, **52**(16), 4905–4921.

Lu, S.Y., & Fu, K.S. 1978. A syntactic approach to texture analysis. *Computer Graphics and Image Processing*, **7**(3), 303–330.

Ma, F., Bajger, M., Slavotinek, J., & Bottema, M. 2006. Validation of graph theoretic segmentation of the pectoral muscle. *Pages 642–649 of: the 8th International Workshop on Digital Mammography.*

Ma, F., Bajger, M., Slavotinek, J.P., & Bottema, M.J. 2007. Two graph theory based methods for identifying the pectoral muscle in mammograms. *Pattern Recognition*, **40**(9), 2592–2602.

Ma, Y., Tay, P.C., Adams, R.D., & Zhang, J.Z. 2010. A novel shape feature to classify microcalcifications. *Pages 2265–2268 of: the 17th IEEE International Conference on Image Processing.*

Mac Parthaláin, N., & Shen, Q. 2009. Exploring the boundary region of tolerance rough sets for feature selection. *Pattern Recognition*, **42**(5), 655–667.

Mac Parthaláin, N., Jensen, R., Shen, Q., & Zwiggelaar, R. 2010. Fuzzy-rough approaches for mammographic risk analysis. *Intelligent Data Analysis*, **14**(2), 225–244.

Mac Queen, J. 1967. Some methods for classification and analysis of multivariate observations. *Pages 281–297 of: the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1.

Malik, J., & Perona, P. 1990. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America, Series A*, **7**(5), 923–932.

Malladi, R., Sethian, J.A., & Vemuri, B.C. 1995. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), 158–175.

Manduca, A., Carston, M.J., Heine, J.J., Scott, C.G., Pankratz, V.S., Brandt, K.R., Sellers, T.A., Vachon, C.M., & Cerhan, J.R. 2009. Texture features from mammographic images and risk of breast cancer. *Cancer Epidemiology Biomarkers & Prevention*, **18**(3), 837–845.

Marčelja, S. 1980. Mathematical description of the responses of simple cortical cells. *Journal of Optical Society of America*, **70**(11), 1297–1300.

Marias, K., Petroudi, S., English, R., Adams, R., & Brady, M. 2002. Subjective and computer-based characterisation of mammographic patterns. *Pages 552–557 of: the 6th Internation Workshop on Digital Mammography.*

Marias, K., Behrenbruch, C., Highnam, R., Parbhoo, S., Seifalian, A., & Brady, M. 2004. A mammographic image analysis method to detect and measure changes in breast density. *European Journal of Radiology*, **52**(3), 276–282.

Marr, D. 1982. *Vison: A computational investigation into the human representation and processing of visual information.* San Francisco: Freeman.

Marr, D., & Hildreth, E. 1980. Theory of edge detection. *Pages 187–217 of: the Royal Society of London. Series B. Biological Sciences*, vol. 207.

Martí, R., Oliver, A., Raba, D., & Freixenet, J. 2007. Breast skin-line segmentation using contour growing. *Pages 564–571 of: the 3rd Iberian Conference on Pattern Recognition and Image Analysis.*

Martin, K.E., Helvie, M.A., Zhou, C., Roubidoux, M.A., Bailey, J.E., Paramagul, C., Blane, C.E., Klein, K.A., Sonnad, S.S., & Chan, H.P. 2006. Mammographic Density Measured with Quantitative Computer-aided Method: Comparison with Radiologists' Estimates and BI-RADS Categories. *Radiology*, **240**(3), 656–665.

Masek, M., Attikiouzel, Y., & deSilva, C.J.S. 2000. Skin-air interface extraction from mammograms using an automatic local thresholding algorithm. *Pages 204–206 of: the 15th Biennial International Conference Biosignal.*

Maskarinec, G., Pagano, I., Lurie, G., Wilkens, L.R., & Kolonel, L.N. 2005. Mammographic Density and Breast Cancer Risk. *American Journal of Epidemiology*, **162**(8), 743.

McCormack, V.A., & dos Santos Silva, I. 2006. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiology Biomarkers & Prevention*, **15**(6), 1159–1169.

McInerney, T., & Terzopoulos, D. 1995. A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis. *Computerized Medical Imaging and Graphics*, **19**(1), 69–83.

McInerney, T., & Terzopoulos, D. 1996. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, **1**(2), 91–108.

Méndez, A.J., Tahoces, P.G., Lado, M.J., Souto, M., Correa, J.L., & Vidal, J.J. 1996. Automatic detection of breast border and nipple in digital mammograms. *Computer Methods and Programs in Biomedicine*, **49**(3), 253–262.

Mikolajczyk, K., & Schmid, C. 2004. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, **60**(1), 63–86.

Miller, P., & Astley, S. 1992. Classification of breast tissue by texture analysis. *Image and Vision Computing*, **10**(5), 277–282.

Mirmehdi, M., Xie, X., & Suri, J.S. 2009. *Handbook of texture analysis*. Imperial College Press.

Morton, A., Chan, H., & Goodsitt, M. 1996. Automated model-guided breast segmentation algorithm. *Medical Physics*, **23**(6), 1107–1108.

Moss, S.M., Cuckle, H., Evans, A., Johns, L., Waller, M., Bobrow, L., *et al.* 2006. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *The Lancet*, **368**(9552), 2053–2060.

Muhimmah, I., & Zwiggelaar, R. 2006. Mammographic density classification using multiresolution histogram information. *In: the 5th International IEEE EMBS Special Topic Conference on Information Technology in Biomedicine*.

Murphy, K., van Ginneken, B., Schilham, A.M., de Hoop, B.J., Gietema, H.A., & Prokop, M. 2009. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Medical Image Analysis*, **13**(5), 757–770.

National Comprehensive Cancer Network. 2012. NCCN Clinical Practice Guidelines in Oncology: Breast Cancer Screening and Diagnosis.

Nie, K., Chang, D., Chen, J.H., Hsu, C.C., Nalcioglu, O., & Su, M.Y. 2010. Quantitative analysis of breast parenchymal patterns using 3D fibroglandular tissues segmented based on MRI. *Medical Physics*, **37**, 217–226.

Nishikawa, R.M. 2007. Current status and future directions of computer-aided diagnosis in mammography. *Computerized Medical Imaging and Graphics*, **31**(4), 224–235.

Nyström, L., Wall, S., Rutqvist, L.E., Lindgren, A., Lindqvist, M., Rydén, S., Andersson, I., Bjurstam, N., Fagerberg, G., Frisell, J., Tabár, L., & Larsson, L.G. 1993. Breast cancer screening with mammography: overview of Swedish randomised trials. *The Lancet*, **341**(8851), 973–978.

Obenauer, S., Luftner-Nagel, S., von Heyden, D., Munzel, U., Baum, F., & Grabbe, E. 2002. Screen film vs full-field digital mammography: image quality, detectability and characterization of lesions. *European Radiology*, **12**, 1697–1702.

Office for National Statistics. 2012. *Cancer incidence and mortality in the UK, 2007-2009*. Statistical Bulletin.

Ojala, T., Näppi, J., & Nevalainen, O. 2001. Accurate segmentation of the breast region from digitized mammograms. *Computerized Medical Imaging and Graphics*, **25**(1), 47–59.

Ojala, T., Pietikäinen, M., & Mäenpää, T. 2002. Multiresolution grey-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(7), 971–987.

Oliver, A., Freixenet, J., & Zwiggelaar, R. 2005a. Automatic Classification of Breast Density. *Pages 1258–1261 of: the 12th IEEE International Conference on Image Processing*, vol. 2.

Oliver, A., Freixenet, J., Bosch, A., Raba, D., & Zwiggelaar, R. 2005b. Automatic Classification of Breast Tissue. *Pages 431–438 of: the 2nd Iberian Conference on Pattern Recognition and Image Analysis*.

Oliver, A., Freixenet, J., Martí, R., & Zwiggelaar, R. 2006. A comparison of breast tissue classification techniques. *Pages 872–879 of: the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention*.

Oliver, A., Freixenet, J., Martí, R., Pont, J., Pérez, E., Denton, E.R.E., & Zwiggelaar, R. 2008. A novel breast tissue density classification methodology. *IEEE Transactions on Information Technology in Biomedicine*, **12**(1), 55–65.

Oliver, A., Lladó, X., Pérez, E., Pont, J., Denton, E.R.E., Freixenet, J., & Martí, J. 2010. A statistical approach for breast density segmentation. *Journal of Digital Imaging*, **23**(5), 527–537.

Oliver, A., Torrent, A., Lladó, X., Tortajada, M., Tortajada, L., Sentís, M., Freixenet, J., & Zwiggelaar, R. 2012. Automatic microcalcification and cluster detection for digital and digitised mammograms. *Knowledge-Based Systems*, **28**, 68–75.

Ooms, E.A., Zonderland, H.M., Eijkemans, M.J.C., Kriege, M., Mahdavian Delavary, B., Burger, C.W., & Ansink, A.C. 2007. Mammography: Interobserver variability in breast density assessment. *The Breast*, **16**(6), 568–576.

Otsu, N. 1975. A threshold selection method from gray-level histograms. *Automatica*, **11**(285-296), 23–27.

Pal, N.R., & Pal, S.K. 1993. A review on image segmentation techniques. *Pattern Recognition*, **26**(9), 1277–1294.

Papadopoulos, A., Fotiadis, D.I., & Likas, A. 2005. Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. *Artificial Intelligence in Medicine*, **34**(2), 141–150.

Pawluczyk, O., Augustine, B.J., Yaffe, M.J., Rico, D., Yang, J., Mawdsley, G.E., & Boyd, N.F. 2003. A volumetric method for estimation of breast density on digitized screen-film mammograms. *Medical Physics*, **30**, 352–364.

Pentland, A.P. 1984. Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**(6), 661–674.

Petroudi, S., & Brady, M. 2006. Breast density segmentation using texture. *Pages 609–615 of: the 8th International Workshop on Digital Mammography.*

Petroudi, S., Kadir, T., & Brady, M. 2003. Automatic classification of mammographic parenchymal patterns: A statistical approach. *Pages 798–801 of: the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*

Pham, D.L., & Prince, J.L. 1998. An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities. *Pattern Recognition Letters*, **20**, 57–68.

Pham, D.L., Xu, C., & Prince, J.L. 2000. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, **2**(1), 315–337.

Pisano, E.D., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J.K., Acharyya, S., Conant, E.F., Fajardo, L.L., Bassett, L., D'Orsi, C., *et al.* 2005. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, **353**(17), 1773–1783.

Pisano, E.D., Hendrick, R.E., Yaffe, M.J., Baum, J.K., Acharyya, S., Cormack, J.B., Hanna, L.A., Conant, E.F., Fajardo, L.L., Bassett, L.W., *et al.* 2008. Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST. *Radiology*, **246**(2), 376–383.

Raba, D., Oliver, A., Martí, J., Peracaula, M., & Espunya, J. 2005. Breast segmentation with pectoral muscle suppression on digital mammograms. *Pages 471–478 of: the 2nd Iberian Conference on Pattern Recognition and Image Analysis.*

Rana, R.S., Jiang, Y., Schmidt, R.A., Nishikawa, R.M., & Liu, B. 2007. Independent evaluation of computer classification of malignant and benign calcifications in full-field digital mammograms. *Academic Radiology*, **14**(3), 363–370.

Rasmussen, C.E. 2000. The infinite Gaussian mixture model. *Advances in Neural Information Processing Systems*, **12**, 554–560.

Raundahl, J., Loog, M., & Nielsen, M. 2006a. Mammographic density measured as changes in tissue structure caused by HRT. *Pages 141–148 of: SPIE Medical Imaging: Image Processing*, vol. 6144.

Raundahl, J., Loog, M., & Nielsen, M. 2006b. Understanding Hessian-based density scoring. *Pages 447–452 of: the 8th International Workshop on Digital Mammography.*

Raundahl, J., Loog, M., Pettersen, P., Tanko, L.B., & Nielsen, M. 2008. Automated effect-specific mammographic pattern measures. *IEEE Transactions on Medical Imaging*, **27**(8), 1054–1060.

Reddick, W.E., Glass, J.O., Cook, E.N., Elkin, T.D., & Deaton, R.J. 1997. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on Medical Imaging*, **16**(6), 911–918.

Redner, R.A., & Walker, H.F. 1984. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, **26**(2), 195–239.

Reed, T.R., & Wechsler, H. 1990. Segmentation of textured images and Gestalt organization using spatial/spatial-frequency representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(1), 1–12.

Ren, J. 2012. ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. *Knowledge-Based Systems*, **26**, 144–153.

Ren, J., Wang, D., & Jiang, J. 2011. Effective recognition of MCCs in mammograms using an improved neural classifier. *Engineering Applications of Artificial Intelligence*, **24**, 638–645.

Rioul, O., & Vetterli, M. 1991. Wavelets and signal processing. *IEEE Signal Processing Magazine*, **8**(4), 14–38.

Ripley, B.D. 1996. *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Rish, I. 2001. An empirical study of the naive Bayes classifier. *Pages 41–46 of: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, vol. 3.

Rissanen, J. 1978. Modeling by shortest data description. *Automatica*, **14**(5), 465–471.

Saftlas, A.F., Wolfe, J.N., Hoover, R.N., Brinton, L.A., Schairer, C., Salane, M., & Szklo, M. 1989. Mammographic parenchymal patterns as indicators of breast cancer risk. *American Journal of Epidemiology*, **129**(3), 518–526.

Saha, P.K., Udupa, J.K., Conant, E.F., Chakraborty, D.P., & Sullivan, D. 2001. Breast tissue density quantification via digitized mammograms. *IEEE Transactions on Medical Imaging*, **20**(8), 792–803.

Sahoo, P.K., Soltani, S., Wong, A.K.C., & Chen, Y.C. 1988. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, **41**(2), 233–260.

Sarkar, M. 2007. Fuzzy-rough nearest neighbor algorithms in classification. *Fuzzy Sets and Systems*, **158**(19), 2134–2152.

Shao, Y.Z., Liu, L.Z., Bie, M.J., Li, C., Wu, Y., Xie, X., & Li, L. 2011. Characterizing the clustered microcalcifications on mammograms to predict the pathological classification and grading: A mathematical modeling approach. *Journal of Digital Imaging*, **24**, 764–771.

Shen, L., Rangayyan, R.M., & Desautels, J.E.L. 1994. Application of shape analysis to mammographic calcifications. *IEEE Transactions on Medical Imaging*, **13**(2), 263–274.

Sickles, E.A. 1986. Breast calcifications: Mammographic evaluation. *Radiology*, **160**(2), 289–293.

Sivaramakrishna, R., Obuchowski, N.A., Chilcote, W.A., & Powell, K.A. 2001. Automatic segmentation of mammographic density. *Academic Radiology*, **8**(3), 250–256.

Skaane, P., Hofvind, S., & Skjennald, A. 2007. Randomized Trial of Screen-Film versus Full-Field Digital Mammography with Soft-Copy Reading in Population-based Screening Program: Follow-up and Final Results of Oslo II Study. *Radiology*, **244**(3), 708–717.

Soille, P. 2003. *Morphological image analysis: Principles and applications*. Second edn. Heidelberg: Springer-Verlag.

Soltanian-Zadeh, H., Rafiee-Rad, F., & Pourabdollah-Nejad, D.S. 2004. Comparison of Multiwavelet, Wavelet, Haralick, and shape features for microcalcification classification in mammograms. *Pattern Recognition*, **37**(10), 1973–1986.

Sonka, M., Hlavac, V., & Boyle, R. 2007. *Image processing, analysis, and machine vision*. 3 edn. Toronto: Thomson Learning.

Stone, J., Warren, R.M.L., Pinney, E., Warwick, J., & Cuzick, J. 2009. Determinants of percentage and area measures of mammographic density. *American Journal of Epidemiology*, **170**(12), 1571–1578.

Subashini, T.S., Ramalingam, V., & Palanivel, S. 2010. Automated assessment of breast tissue density in digital mammograms. *Computer Vision and Image Understanding*, **114**(1), 33–43.

Suckling, J., Parker, J., Dance, D.R., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S.L., Taylor, P., Betal, D., & Savage, J. 1994. The Mammographic Image Analysis Society digital mammogram database. *Pages 375–378 of: Excerpta Medica. International Congress Series 1069*.

Sun, X., Qian, W., & Song, D. 2004. Ipsilateral-mammogram computer-aided detection of breast cancer. *Computerized Medical Imaging and Graphics*, **28**(3), 151–158.

Szilagyi, L., Benyo, Z., Szilagyi, S.M., & Adam, H.S. 2003. MR brain image segmentation using an enhanced fuzzy c-means algorithm. *Pages 724–726 of: the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.

Szilágyi, L., Szilágyi, S., & Benyó, Z. 2007. A modified fuzzy c-means algorithm for MR brain image segmentation. *Pages 866–877 of: the 4th International Conference on Image Analysis and Recognition.*

Tabár, L., Yen, M.F., Vitak, B., Chen, H.H.T., Smith, R.A., & Duffy, S.W. 2003. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *The Lancet*, **361**(9367), 1405–1410.

Tabár, L., Tot, T., & Dean, P.B. 2005. *Breast cancer: the art and science of early detection with mammography: perception, interpretation, histopathologic correlation.* George Thieme Verlag.

Tabár, L., Vitak, B., Chen, T.H.H., Yen, A.M.F., Cohen, A., Tot, T., Chiu, S.Y.H., Chen, S.L.S., Fann, J.C.Y., Rosell, J., Fohlin, H., Smith, R.A., & Duffy, S.W. 2011. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*, **260**(3), 658–663.

Tahoces, P.G., Correa, J., Soutos, M., Gómez, L., & Vidal, J.J. 1995. Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns. *Physics in Medicine and Biology*, **40**, 103–117.

Thompson, A.C. 1996. *Minkowski Geometry, volume 63 of Encyclopedia of Mathematics and its Applications.* Cambridge University Press.

Torrent, A., Bardera, A., Oliver, A., Freixenet, J., Boada, I., Feixes, M., Martí, R., Lladó, X., Pont, J., Pérez, E., *et al.* 2008. Breast density segmentation: a comparison of clustering and region based techniques. *Pages 9–16 of: the 9th International Workshop on Digital Mammography.*

Tuceryan, M. 1994. Moment-based texture segmentation. *Pattern Recognition Letters*, **15**(7), 659–668.

Tuceryan, M., & Jain, A.K. 1990. Texture segmentation using voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(2), 211–216.

Tuceryan, M., & Jain, A.K. 1993. Texture analysis. *Chap. 2, pages 207–248 of:* C. H. Chen, L. F. Pau, P. S. P. Wang (ed), *Handbook of pattern recognition and computer vision.* World Scientific Publishing Co.

Tzikopoulos, S.D., Mavroforakis, M.E., Georgiou, H.V., Dimitropoulos, N., & Theodoridis, S. 2011. A fully automated scheme for mammographic segmen-

tation and classification based on breast density and asymmetry. *Computer Methods and Programs in Biomedicine*, **102**(1), 47–63.

Udupa, J.K., & Samarasekera, S. 1996. Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. *Graphical Models and Image Processing*, **58**(3), 246–261.

Unser, M. 1995. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*, **4**(11), 1549–1560.

Unser, M., & Eden, M. 1990. Nonlinear operators for improving texture segmentation based on features extracted by spatial filtering. *IEEE Transactions on Systems, Man and Cybernetics*, **20**(4), 804–815.

Ursin, G., Astrahan, M.A., Salane, M., Parisky, Y.R., Pearce, J.G., Daniels, J.R., Pike, M.C., & Spicer, D.V. 1998. The detection of changes in mammographic densities. *Cancer Epidemiology Biomarkers & Prevention*, **7**(1), 43–47.

Vachon, C.M., van Gils, C.H., Sellers, T.A., Ghosh, K., Pruthi, S., Brandt, K.R., & Pankratz, V.S. 2007. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Research*, **9**(6), 217.

van Engeland, S., Snoeren, P.R., Huisman, H., Boetes, C., & Karssemeijer, N. 2006. Volumetric breast density estimation from full-field digital mammograms. *IEEE Transactions on Medical Imaging*, **25**(3), 273–282.

Varma, M., & Zisserman, A. 2005. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, **62**(1), 61–81.

Varma, M., & Zisserman, A. 2009. A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**(11), 2032–2047.

Vigeland, E., Klaasen, H., Klingen, T.A., Hofvind, S., & Skaane, P. 2008. Full-field digital mammography compared to screen film mammography in the prevalent round of a population-based screening programme: the Vestfold County Study. *European Radiology*, **18**(1), 183–191.

Vrooman, H.A., Cocosco, C.A., van der Lijn, F., Stokking, R., Ikram, M.A., Vernooij, M.W., Breteler, M., & Niessen, W.J. 2007. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage*, **37**(1), 71–81.

Warfield, S.K., Zou, K.H., & Wells, W.M. 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, **23**(7), 903–921.

Warren Burhenne, L.J., Wood, S.A., D'Orsi, C.J., Feig, S.A., Kopans, D.B., O'Shaughnessy, K.F., Sickles, E.A., Tabár, L., Vyborny, C.J., & Castellino, R.A. 2000. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology*, **215**(2), 554–562.

Wei, J., Chan, H.P., Helvie, M.A., Roubidoux, M.A., Sahiner, B., Hadjiiski, L.M., Zhou, C., Paquerault, S., Chenevert, T., & Goodsitt, M.M. 2004. Correlation between mammographic density and volumetric fibroglandular tissue estimated on breast MR images. *Medical Physics*, **31**(4), 933–942.

Wei, J., Chan, H.P., Wu, Y.T., Zhou, C., Helvie, M.A., Tsodikov, A., Hadjiiski, L.M., & Sahiner, B. 2011. Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study. *Radiology*, **260**(1), 42–49.

Wei, J., Chan, H.P., Lu, Y., Hadjiiski, L., Zhou, C., & Helvie, M. 2012. Breast Parenchymal Pattern (BPP) Analysis: Comparison of Digital Mammograms and Breast Tomosynthesis. *Breast Imaging*, 514–520.

Wei, L., Yang, Y., & Nishikawa, R.M. 2009. Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recognition*, **42**(6), 1126–1132.

Weszka, J.S. 1978. A survey of threshold selection techniques. *Computer Graphics and Image Processing*, **7**(2), 259–265.

Wirth, M., Fraschini, M., & Lyon, J. 2004. Contrast enhancement of microcalcifications in mammograms using morphological enhancement and non-flat structuring elements. *Pages 134–139 of: Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems.*

Wirth, M., Nikitenko, D., & Lyon, J. 2005. Segmentation of the breast region in mammograms using a rule-based fuzzy reasoning algorithm. *International Journal on Graphics, Vision and Image Processing*, **5**(2), 45–54.

Wirth, M.A., & Stapinski, A. 2003. Segmentation of the breast region in mammograms using active contours. *Pages 1995–2006 of: Visual Communications and Image Processing*, vol. 5150.

Wolfe, J.N. 1976a. Breast patterns as an index of risk for developing breast cancer. *American Journal of Roentgenology*, **126**(6), 1130–1137.

Wolfe, J.N. 1976b. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, **37**(5), 2486–2492.

Wong, A.K.C., & Sahoo, P.K. 1989. A gray-level threshold selection method based on maximum entropy principle. *IEEE Transactions on Systems, Man and Cybernetics*, **19**(4), 866–871.

Xia, G.S., Delon, J., & Gousseau, Y. 2010. Shape-based invariant texture indexing. *International Journal of Computer Vision*, **88**(3), 382–403.

Xu, C., Pham, D.L., & Prince, J.L. 2000. Image segmentation using deformable models. *Pages 129–174 of:* Fitzpatrick, J. M., & Sonka, M. (eds), *Handbook of Medical Imaging*. Medical Image Processing and Analysis, vol. 2. Bellingham, WA: SPIE.

Yaffe, M.J. 2008. Measurement of mammographic density. *Breast Cancer Research*, **10**(3).

Yin, F.F., Giger, M.L., Doi, K., Vyborny, C.J., & Schmidt, R.A. 1994. Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique. *Medical Physics*, **21**(3), 445–452.

Yuille, A.L., & Poggio, T.A. 1986. Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(1), 15–25.

Zadeh, L.A. 1965. Fuzzy sets. *Information and Control*, **8**(3), 338–353.

Zhang, J., Marszalek, M., Lazebnik, S., & Schmid, C. 2007. Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision*, **73**(2), 213–238.

Zhang, L., Hoffman, E.A., & Reinhardt, J.M. 2006. Atlas-driven lung lobe segmentation in volumetric X-ray CT images. *IEEE Transactions on Medical Imaging*, **25**(1), 1–16.

Zhang, Y., Brady, M., & Smith, S. 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, **20**(1), 45–57.

Zhou, C., Chan, H.P., Petrick, N., Helvie, M.A., Goodsitt, M.M., Sahiner, B., & Hadjiiski, L.M. 2001. Computerized image analysis: Estimation of breast density on mammograms. *Medical Physics*, **28**(6), 1056–1069.

Zhou, C., Chan, H.P., Paramagul, C., Roubidoux, M.A., Sahiner, B., Hadjiiski, L.M., & Petrick, N. 2004. Computerized nipple identification for multiple image analysis in computer-aided diagnosis. *Medical Physics*, **31**(10), 2871–2882.

Zucker, S.W. 1976. Toward a model of texture. *Computer Graphics and Image Processing*, **5**(2), 190–202.

Zwiggelaar, R. 2010. Local greylevel appearance histogram based texture segmentation. *Pages 175–182 of: the 10th International Workshop on Digital Mammography.*

Zwiggelaar, R., & Denton, E. 2006. Texture based segmentation. *Pages 433–440 of: the 8th International Workshop on Digital Mammography.*

Zwiggelaar, R., Blot, L., Raba, D., & Denton, E. 2003. Set-permutation-occurrence matrix based texture segmentation. *Pages 1099–1107 of: the 1st Iberian Conference on Pattern Recognition and Image Analysis.* Springer.

Zwiggelaar, R., Astley, S.M., Boggis, C.R.M., & Taylor, C.J. 2004. Linear structures in mammographic images: detection and classification. *IEEE Transactions on Medical Imaging*, **23**(9), 1077–1086.

Zwiggelaar, R., Muhimmah, I., & Denton, E.R.E. 2005. Mammographic density classification based on statistical grey-level histogram modeling. *Pages 183–186 of: the 9th Medical Image Understanding and Analysis.*